5-2015

The Utility of Local Norms in the Identification of Students with Learning Disabilities in Reading

Brian C. Malliett
Olivet Nazarene University, bcmalliett@gmail.com

THE UTILITY OF LOCAL NORMS IN THE IDENTIFICATION OF STUDENTS WITH LEARNING DISABILITIES IN READING

by

Brian C. Malliett

Dissertation

Submitted to the Faculty of

Olivet Nazarene University

School of Graduate and Continuing Studies

in Partial Fulfillment of the Requirements for

the Degree of

Doctor of Education

in

Ethical Leadership

May 2015

THE UTILITY OF LOCAL NORMS IN THE IDENTIFICATION OF STUDENTS WITH LEARNING DISABILITIES IN READING

by

Brian C. Malliett

Dissertation

Dissertation Advisor	4/18/15 Date
Dissertation Reader	4-18-15 Date
Dissertation Coordinator	4-18-15 Date
Program Director	4-18-15 Date
Vice President for Academic Affairs	4-18-15 Date

ACKNOWLEDGMENTS

I have learned to reach out and ask for help from an army of support. A legion of advocates supported me through the dissertation *process*. The *process* was more than a terminal degree for one person. This academic challenge of one led to the realization that illuminated the intricate network of those who were willing and able to support each other. These individuals pushed me harder and longer than I ever thought. It is my hope that I can, in return, give back this strength to others through my professional and personal interactions with others. My sincerest appreciation and gratitude goes out to my family, advisor/s, instructors, classmates, colleagues, and friends. I would love to name them, but I do believe in confidentiality and the possibility that I could leave someone out of the list of many. I will mention Cohort VII as my surrogate family. Unfortunately, our time together enduring through the volume of the workload and trials of education has come to an end. Next time we see each other will be in the fascinating world we left behind three years ago. Those individuals have become my newest role models. But most importantly, the care and concern of others that made it possible for one person's long-term goal to come into fruition

ABSTRACT

Third grade students were administered two universal screening reading measures at the beginning and end of the school year. In an effort to limit the misidentification of reading disabilities (RD) in lower socioeconomic community populations of the school district, local norms were established. The local and national norms were compared using the 10th percentile, or -1.27 *Z*-score, to identify discrepant scores. A diminished classification of RD was evidenced when local norms were utilized in contrast to the use of national norms. Locally-normed slope scores in conjunction with the benchmark testing scores showed a moderate level of correlation than when compared to the scores or slope coefficients in isolation. The utilization of local norms reduced the number of classifications for RD, but did not consistently identify the children who were diagnosed with a Specific Learning Disability (SLD). The identification of statistically significant discrepancies of the Progress Monitoring (PM) scores provided support in the practice for the assessment of *SLD* when used jointly with slope scores. Overall, the use of *Z*-scores offered a reliable means for the comparison of the various PM measure scores.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The investigation addressed the standardization of interval assessment data generated based on Individuals with Disabilities Education Improvement Act of 2004 (IDEA) evaluation requirements for learning disabilities (United States Department of Education, 2013). The State of Illinois recognized 108,386 of the total of 2,087,628 students with identified learning disabilities during the school year 2011-12 (Illinois State Board of Education, 2012a). The percentage of students with Special Education services for Specific Learning Disabilities (SLD) was 5.1%. McNamara and Willoughby (2010) cited a national range of 4 to 5% as consistent with recent prevalence research. Throughout the state, the frequencies of eligible students with learning disabilities ranged considerably. The misidentification of learning disabilities, especially with diverse lower socioeconomic students, has required the Illinois state education system to monitor the number of children found eligible for learning disabilities.

School districts across the state developed various evaluation policies in order to assess and diagnose SLD. However, the evaluation of SLD varied between the school districts due to curriculum differences and local population demographics across the state. School districts across the state adopted and/or developed their own procedures in order to identify students with SLD based on their local population's academic performance (Illinois State Board of Education, 2012a). Curriculum-based performance

scores were typically used to evaluate the identification of learning disabilities based on legislative mandates (United States Department of Education, 2013; Zirkel, 2013). The system of monitoring student growth over time, the assessment of poor academic growth, and the diagnosis of SLD was determined by the school districts.

The problem resided with the difficulties associated with the numerous procedures that school districts utilized to assess SLD using Progress Monitoring (PM)data, measuring student outcomes to interventions, and the norms used to determined discrepant scores (Utley, Oiakor, & Bakkon, 2011; Wedl, 2005). Several measures provided the means to assess reading skills based on the students' response to the schools' curricula. The universal screening or benchmark PM measures provided a variety of scores, which included both raw and scale scores. Universal screening tests were used at specific times of the school year. These evaluations were administered to students in the fall when they started school, winter at midterm, and spring, or the end of the school year. These scores were reflected as single entities, rather than a cohesive set of data points that were compared.

The benchmark scores were based on national norms that tended not to reflect the academic performance of the school districts with children who have a higher incidence of poverty, language, and ethnic differences (Skiba, Poloni-Staudinger, Simmons, Feggins-Azziz, & Chung, 2005; Utley, 2011). Standardizing the scores through local norming provided the practice for increased consistency of the decisions made by the school teams to identify children with SLD, especially when multiple measures were utilized (Gresham, 2009; Stecker, Lempke, & Foegan, 2008; Van Der Heyden, 2011; Yeo, Fearington, & Christ, 2011).

The data generated from PM measures that were normed nationally do not reflect local populations with higher proportions of minorities who are impacted with lower socioeconomic conditions. Often, minority children from lower socioeconomic families and communities faced life styles that interfered with educational expectations (Yeo et al., 2011). These conditions reflected many factors related to transiency, truancy, language, homelessness, safety, and needed medical care. The effects of these factors can negatively impact academic performance (Utley et al., 2011). Thus, scores on PM testing were impacted as a result. These factors did not mean that the children presented with learning disabilities, but rather the life conditions interfered with the educational process. When a child was enduring these conditions, the focus was on the requirements of life and not necessarily education (Sullivan et al., 2009). When the national norms were used to assess minority student progress, the results often misrepresented local lower socioeconomic students by the misidentification of children with potential learning disabilities. Using local norms was an attempt to minimize the over-identification of students with learning disabilities and provided the information needed to support the children who require interventions based on the local curriculum.

The PM results came in a variety of reported scores. These scores ranged from scale to raw scores that had little or no inherent meaning to the casual user. When these various scores were used, they were often used in isolation, rather than in conjunction to support decision making. In order to use the scores synergistically, it made sense to standardize the scores in order to use them for assessment purposes. Converting these various reported PM results to *Z*-scores could help to simplify the analysis of scores in order to utilize them in a collective manner for assessment. Furthermore, the conversion

of the scores led the development of local norming, which supported curriculum decision making for school district populations impacted by lower socioeconomic and difficult community circumstances. The *Z*-score system was used for instructional intervention monitoring and for the identification and diagnosis of SLD (Stewart & Silberglitt, 2008). The intention of the study was to accurately identify students with potential learning disabilities who presented with discrepant reading-based locally-normed PM scores (Francis, Fletcher, Steubing, & Lyon, 2005; Reynolds & Shaywitz, 2009).

Statement of the Problem

The problem resided with the determination of statistically significant score discrepancies and the lack of student progress based on the PM measures used by the SLD teams (Yeo et al., 2011). The determination of the discrepancy or lack of progress has not been extensively researched previously with the standardized use of local norms (Burns, 2007). Standardizing the PM scores using *Z*-scores may provide the measure of assurance that decisions about student achievement or lack of progress were consistently and accurately assessed (Ball & Christ, 2012; Compton, et al., 2010; Fuchs & Fuchs, 2006). Standardized testing results supported comparisons across similar scales while supporting IDEA 2004 evaluation criteria for the evaluation of SLD. The ability to compare other scales would provide the assessment teams with data analysis information and improved consistency with decision making (Shapiro et al., 2011).

The purpose of the current study was to determine reliable and valid academic performance discrepancies based on various locally-normed benchmark reading-based PM scores along with slope coefficients in order to standardize eligibility criteria for students suspected of learning disabilities in reading (Marston, & Magnusson, 1988). The

Z-score was used for the standardization of the data. Local norming was developed to better assess the conditions that can negatively impact the learning needs of the children in the study population. Local norms provided a diminished bias on a local population instead of using national norms (Hosp, Hosp, & Dole, 2011; Stewart & Silberglitt, 2008). The current study was undertaken to minimize the misidentification of SLD children who were impacted by the conditions related to lower socioeconomic communities and proposed a valuable set of guidelines and safeguards to minimize the misidentification of SLD in reading (Hale, et al., 2010).

Background

Changes in IDEA 2004 directly influenced the determination of learning disabilities. Legislators mandated the school assessment teams to collect data that would support or refute a lack of student response to research-based interventions. The updated federal education law that addressed student disabilities reflected a change to the manner that SLD was evaluated. One of the stipulation changes dropped by the legislators of IDEA 2004 was the sole use of the ability - achievement discrepancy model as the primary diagnostic criteria for the identification of SLD (Shinn, 2007; Zhang, 2007; Zirkel & Rose 2009). Instead, a process of researched-based interventions must be attempted first to assess whether or not the student presented with academic performance deficits or may have a potential learning disability.

The direct reflection approach to norming benchmark scores, based on the local population provided valuable information to ensure that consistent decisions were made. Local norms reflected the population that was directly impacted by the curriculum (Ball & Christ, 2012). When using local norms, the PM data were analyzed statistically to

reflect the measurement qualities within a given population (Johnson, Jenkins, Petscher, & Catts, 2009; Parker, Vannest, Davis, & Clemens, 2012).

Progress monitoring provided the opportunity for schools to monitor and identify students who were struggling. These students with academic limitations were provided differentiated instruction in the deficit areas (Fuchs, Fuchs, & Hollenbeck, 2007). When the student did not respond to the directed interventions in the area of the identified deficit, PM measures provided the means to assess academic performance over time. Instead of a one-time battery evaluation of the student to assess for a learning disability, the student was administered several PM instruments at periodic intervals along with research-based interventions to assess for academic progress.

The legislators of IDEA 2004 stipulated that the use of academic monitoring assessment data was essential in the process of identifying SLD eligibility requirements and exclusionary criteria. A student's scores required discrepant scores when compared to peer test results. A student's poor response to instruction was necessary to consider learning disabilities as evidenced relative to the research-based interventions. The PM performance of the student that indicated a poor response was based on a comparison to the mean scores of the local grade level and seasonal scores of the given population (Shinn, 2007). The lawmakers who drafted IDEA 2004 specified that the documentation of student response to research-based interventions must be identified before finding a student eligible for a learning disability classification found in the IDEA 2004 Regulations, § 300.307(a)(2) (United States Department of Education, 2013).

IDEA 2004 legislation had created a new consideration of PM measures as the source of the evaluation data that addressed a child's lack of academic growth. The

adoption of the federal legislation by the states that addressed the laws governing the identification of SLD has led to the debate of when a student was identified as SLD. In practice the question was, at what point in time or score profile could an assessor adequately make a case for SLD? Children who presented with academic deficits based on PM data were referred to assessment teams with a wide array of scores and profiles. Another major concern was based on limiting the misidentification of SLD.

Ahearn (2009) addressed the importance of the consistent identification of children who were not responding to interventions. In order to develop a system to compare the various scores of the benchmark measures, a stable method for the identification of children with potential learning disabilities was essential (Ball & Christ, 2012; Kavale, 2005). The data generated by the various PM measures were standardized using *Z*-scores as a means to compare progress across the measures.

In order to determine the presence of a learning disability, certain conditions must have been met. These conditions entailed a research-based general education curriculum, scientifically-based data collected through PM measures from children who were involved in interventions, children who responded positively to the curriculum or interventions, and data from the PM measures that reflected a severe discrepancy or lack of progress based on grade and age-related peers (IDEA 2004; Mellard, McKnight, & Jordan, 2010; Parker et al., 2012; Zirkel & Rose, 2009).

The researched-based curriculum was fundamental in order to ensure efficacy of the instruction. Once the core curriculum was established, the schools were required to monitor the progress of the students with seasonal benchmark testing (Christ, Silberglitt, Yeo, & Cormier, 2010). Seasonal testing refers to the PM testing that was based on

grade-level expectations for reading, math, and writing. The benchmark testing occurred during the fall, winter, and spring of the school year, thus implying seasonal testing. Students who were not making adequate gains, as evidenced by declining or static scores, were identified through universal screening methods. Research-based interventions were considered after all efforts to address the child's academic needs had been attempted at the general education level.

The standardization of the PM measures further provided reliability to decisions that led to SLD eligibility or for continuing the intervention process through general education tier-supported services (Kavale & Spaulding, 2008). Stollar et al. (2008) asserted that the three tiers of intervention were the outcome of the IDEA 2004 legislation to address a child's response to differentiated instruction. The terms *RTI* or *RTI process* and *Multi-Tiered Support Systems* or *MTSS* will be used interchangeably with and for each other throughout the current study.

The first level of the tier system focused on the general education curriculum expectations that were at the forefront of the instructional process, which was also referred to as Tier-1 interventions. Interventions for the children were considered for those who were not responding solely based on general education instruction. These struggling students were identified from their lower scores on the universal screening measures. Tier-2 interventions were determined when the MTSS team identified the students had not responded adequately to the general education instructional and intervention expectations. Additional intensified instruction in a smaller group with the continued support of the general education curriculum was indicated. Smaller instructional groups of four to eight children with a certified instructor constituted a Tier-

2 intervention. The interventions provided included scientific-based intervention strategies. Approximately four to six weeks were allowed to assess whether or not the child responded to the Tier-2 interventions. When a student was not found to have responded favorably to the Tier-2 level of interventions, the next level of intensity was indicated. The third tier level of intervention provided these students with differentiated instruction in a smaller group of one to three children with the support of certified teacher. During that timeframe, increased PM was provided.

Tier-3 interventions typified the highest level of direct supplemental interventions to children who struggled academically. The group size was typically one to three children with a certified instructor. Reading-based software programs were used to supplement the direct instruction services of a certified teacher. Once a child demonstrated a lack of response to the academic interventions, special education eligibility considerations were explored (Jenkins & Terjeson, 2011). At this point, the educational team explored the possibility of a learning disability. The question that required resolution was what scores were considered discrepant?

The importance of the decision of whether further academic interventions were recommended or referral for a special education evaluation was critical (Batsche, Curtis, Dorman, Castillo, & Porter, 2008; Tran, Sanchez, Arellano, & Swanson, 2010). This decision process was especially evidenced at the crucial junction of third graders who were making the transition from *learning to read* to *reading to learn*. At what cut-off point based on a PM scores should students be classified as making or not making progress?

Another factor included the use of local norms instead of national norms. Students in lower socioeconomic communities were often misidentified with learning disabilities disproportionate to middle-class students (Harry & Klinger, 2006). The interpretation of the scale and raw scores, when using the screening instruments, which presented as discrepant along with slope scores that would assume a lack of progress, was essential. Furthermore, to minimize the misidentification of learning disabilities, local data were necessary to compliment the achievement level of the students in the lower socioeconomic communities.

In order to understand the process of classification and diagnosis for SLD, certain questions required further examination.

Research Questions

1. What differences exist in the identification of specific learning disabilities when using local norms classification compared to national norms classification?
2. What resulting cut-off score is determined when utilizing slope based on locally-normed data?
3. How was the accuracy of the identification of specific learning disabilities affected when both performance scores and slope differential were used?

Description of Terms

The vocabulary that described educational constructs, like any other jargon of occupations or specializations, required definition and exposure to acclimate the reader to the catch phrases and acronyms. Academic lingo was used liberally without the benefit of understanding the fundamental constructs behind the terms in the educational

community. Each of the areas of key research were described and defined to help ensure the understanding of the jargon associated with special education research regarding the identification and diagnostics of SLD.

Special Education services were based on the federal laws pertaining to individuals with disabilities. The name of the Public Law was known as the Individuals with Disabilities Education Improvement Act of 2004 or IDEA (United States Department of Education, 2013). Since the initial drafting of Public Law 94-142, the definitions of the various eligibilities have evolved. One of the most debated eligibility's included the operational definition of SLD. The definition for learning disabilities has a long history of attempts to develop a consensus among clinicians and practitioners (Hale et al., 2010). The definition of SLD concern was on the forefront of these changes.

Benchmark Testing. The seasonal interval testing that was used through the RTI process. It was usually administered in 12 to 18 week intervals in accordance with grading periods to mark the academic progress of all students. The scores were used to monitor student progress related to the instructional efficacy of the curriculum provided (Glover & Diperna, 2007).

Curriculum-Based Evaluation. This form of educational assessment provided a direct connection to the student response to the curriculum. Probes were used to assess reading, math, and written language based on the textbooks and materials utilized in the student's daily instruction. These probes yielded raw scores that were compared to their peers and used to make intervention decisions (Deno, Fuchs, Marston, & Shinn, 2001).

Differentiated Instruction. The theory of instruction that addressed content, process, and product to facilitate the different learning needs of students (Bramlett, Cates, Savina, & Lauinger, 2010).

Discrepancy. The difference between the mean score of a given psycho-educational measurement scale of the population and the students score. A discrepancy reflected deficient scores that were at statistically significant levels below the mean population (Francis et al., 2005). Discrepancy was described in terms of standard deviations below the norm. Discrepant results were used in context with students' scores that reflected moderate to severe academic deficits (Dombrowski, Reynolds, & Kamphaus, 2004).

Disproportionality. The evaluation process where minority students, influenced by lower socioeconomic, cultural, and ethnic origins, were over identified with emotional, behavioral, learning, and intellectual disabilities at a greater incidence rate than the mean population (Illinois State Board of Education, 2012b; Skiba et al., 2005; Sullivan & Artilles, 2011; Sullivan et al., 2009).

Exclusionary Criteria. Certain symptoms of various mental health and medical conditions led to lower academic performance, i.e., Attention Deficit-Hyperactivity Disorder, an asthmatic child who was often absent from school, and other medical conditions that limit a child's attendance or participation at school. Other exclusionary factors included truancy, transiency, homelessness, and criminal detention (United States Department of Education, 2013).

Highly qualified teachers. Teachers that met the rigorous state standards in a subject area/s that placed emphasis on quality instruction were classified as highly

qualified instructors. This standard of quality instruction was essential to the efficacy of the RTI Model and a hallmark of No Child Left Behind Act of 2001 legislation for teachers (Chin, & Wong, 2013; Karelitz, Fields, Levy, Martinez-Gudapakkam, & Jablonski, 2011; Kuntz & Eulass, 2009; No Child Left Behind Act, 2001).

Individuals with Disabilities Education Improvement Act 2004. The Public Law known as the Individuals with Disabilities Education Improvement Act of 2004 was initially drafted in the middle 1970s. This was the public law that governed the procedural rights, the eligibility process, and the services offered to students with disabilities that interfered or created adverse effects in a student's academic performance (34 C.F.R. § 300 United States Department of Education, 2013).

Lack of Progress. This term referred to the poor response or lack of response to academic interventions provided over time to children engaged in the interventions process (United States Department of Education, 2013).

Lexile: A framework of scores that are used to help identify the reading ability range of a student. The score corresponds with the individual's ability to comprehend at least 75% of the reading content that was referred to as the targeted reading level. Many books offered this reading level score to assist with the choice of a book that would best reflect the student's reading level. Scores start below 50, which were at a pre-primer reading level, and ranged well above 1000, reflective of college and graduate level reading content (Meta Metrics, 2015).

Local Norms. The standardization and validation of test scores conducted within a local population (Canter, 1995; Shinn, 1988). These community-based norm-yielded

scores are utilized to represent the student scores of a school district and/or individual school (Hadedank, 1995; Stewart & Silberglitt, 2008).

Misidentification of LD: The over- and under-classification of specific learning disabilities involving the children who were impacted by environmental and familial difficulties of lower socioeconomic communities (Skiba et al., 2005). Studies have reflected a negative impact on learning in children who live in impoverished communities (Utley, 2011). The evaluation process that used national norms was misinterpreted due to the use of nationally-normed scores (Utley et al., 2011).

Multi-Tiered Systems of Support. Also referred to a MTSS, this general education support system better characterized the process of student response to the academic intervention tier system (Gersten, 2009).

National Norms. The validation of tests conducted based on samples within the national population. The normalization of scores on a given testing instrument that were compared or contrasted with other normalized testing instruments. The scores reflect a normal distribution (Salkind, 2012).

No Child Left Behind Act of 2001. The focus of the educational legislation addressed the importance of the promotion of student academic achievement. Teachers were expected to continually improve and meet high standards of instructional performance. Education curricula were modeled on research-based instruction practices. The law placed high standards on student performance that yielded annual yearly progress goals. The student performance was monitored through high-stakes achievement testing (No Child Left Behind Act, 2001).

Progress Monitoring. The repeated interval testing that provided data to measure academic performance over time (Shinn, 2007).

Progress Monitoring Measures. The measuring instruments used to monitor progress and academic performance of students over time (Brown-Childsey & Steege, 2010).

Response to Intervention (RTI). RTI was developed as a general education initiative (National Center for Response to Intervention, 2009). It was a process that allowed schools to monitor and provide supports for children who were struggling academically in school. The interventions were monitored closely through interval testing to ensure that interventions were provided with efficacy (Kuntz, & Eulass, 2009).

Specific Learning Disability. The definition of SLD presented with a broad spectrum of deficits that impaired the neurological processing abilities that effected learning centers (Hale, et al., 2010; 34 C.F.R. § 300.7[c][10]I United States Department of Education, 2013). Learning disabilities or disorders were described as a broad spectrum of neurological conditions that impacted learning. Learning disabilities were highly debated and disputed based on the manner in which a student was identified and diagnosed (Scanlon, 2013; Zirkel, 2010). The learning disabilities that were addressed in the current study will focus on learning problems that negatively impacted reading skills.

Slope. The formula used to develop trend-lines of student data procured through PM measures. This provided a graphical illustration of progress or lack of progress based on the slope of the line. An upward trend provided a positive rise and a negative trend provided a negative decline (Tukey, 1977).

Standard Scores: The normalization of scores that aligned with a normal distribution. These scores were compared and contrasted across measures that reflected similar constructs (Salkind, 2012). In the current study, standard scores known as *Z*-Scores were utilized.

Type I Error: This type of error was based on the researcher's rejection of a null hypothesis that was really true (Salkind, 2012). An example would include indicating that a person had a learning disability when he or she in fact did not.

Type II Error: This statistical error occurs when the researcher fails to reject the null hypothesis that was actually false (Salkind, 2012). An example would include indicating that a person did not have a learning disability when in fact he or she did.

Tier System: The system used by most public schools in the United States included three tiers of intervention. The *tiers* provided a progressive level system of increased instructional content and fewer children in the instructional group. The groups were designed to support children with varying levels of academic need. The more deficits evidenced supported an increased level of interventions (Brown-Childsey & Steege, 2010).

Z-scores: A form of standard scores with a mean of 0 and a standard deviation of 1, where a positive *Z*-score indicated scores above the mean and a negative score reflected scores below the mean (Salkind, 2012).

Significance of the Study

The identification of learning disabilities was based on the students who were not making adequate academic gains. The poor academic performance was based on the students' relative age and grade level. The Illinois State Board of Education (2012a)

established grade-level expectations that provided the standards on which to base the curriculum expectations, but did not offer the expectations of how school personnel should interpret student PM data. The discrepancy or lack of progress required was based on the grade-related children. The universal screening scores provided the data that supported the lack of academic performance improvement (Hixson, Christ, & Bradley-Johnson, 2008). The question addressed by the school personnel was focused on the level of statistical significance that was necessary to diagnose the learning disability or continue the intervention process based on the PM data.

The development of local norms was important to determine in order to provide comparisons in context of grade and/or age interpretation of PM data generated by students within a given curriculum (Stewart & Silberglitt, 2008). The norms were typically developed from data that were collected three or more times per school year. This PM data was a reflection of how the children in general were responding to the curriculum, but more importantly, these norms were essential in order to identify the children who presented with academic needs. Local PM norms were used to guide decision-making. These norms were relevant to the students' school district-based instructional curriculum, rather than comparison to national norms (Stewart & Silberglitt, 2008).

Process to Accomplish

By the time the child reached the third grade, the academic expectations of reading exponentially increased. With the prior development of word decoding, reading fluency, vocabulary, and comprehension skills that were expected of the third-grade students based on No Child Left Behind Act of 2001 expectations, students used reading

as a tool for learning new information. When reading skills progressed at a lower rate than those of their peers, there was a need for a system of monitoring assessments that identified children with reading deficits (Aaron, Joshi, Gooden, & Bentum, 2008; Good, Simmons, & Kame'enui, 2001). While seasonal benchmark screening assessed student academic growth, it also helped to ensure that children were identified as needing supplemental supports and interventions in reading. This process of monitoring student progress led to other concerns related to the interpretation of the data generated that included when to evaluate for SLD.

Monitoring the progress of students presented an essential form of measuring how children were responding to the curriculum. The universal screening measures needed to evaluate the students' progress quickly, and then provide essential information in order to be able to monitor student academic performance relative to the local curriculum (Christ, Zopluoglu, Monaghen, & Van Norman, 2013). Although the data guided the intervention decisions for the struggling students, the data was also a source of information that misleads the decision makers (Gersten et al., 2009). Most of the results of the PM measures were based on national norms that did not accurately address the needs of the children in particular socioeconomic and culturally diverse populations.

Local norms utilized by the individual school or school district were used as the comparison baseline. The average or mean was compared to the student scores to assess growth. Local norming provided the data that guided the school's curriculum-based interventions at the whole school and individual levels (Stewart & Silberglitt, 2008). Progress monitoring measures that addressed the academic needs of the children provided the essential eligibility process of SLD based on IDEA 2004 legislation.

The measures used for PM throughout the educational system were based on various content related scales that helped the school teams interpret a student's academic progress. For the current study, the two reading measures used included the *Dynamic Indicators of Basic Early Literacy Skills (DIBELS)* and the *Performance Series*, 9th *Edition Reading* (*PSR*). These measures offered reliable and valid results for monitoring student academic performance in reading. Both of these universal screening measures have been used extensively by schools throughout the United States. The universal screening assessments provided reliable and valid scores for the routine monitoring of student progress in reading. These measures provided the data that were used to show student progress over time was evidenced by the utilization of the slope model and the difference between the student's academic performance when compared to the grade related norms. The utilization of both the slope and the differential scores supported a method to identify students with potential learning disabilities.

The *DIBELS Oral Reading Fluency* (*DORF*) measured reading skills that included fluency and comprehension. The Cronbach's Alpha coefficient was .91, and ranged between .89 and .93 among the three seasonal administrations of the *DORF* (Shaw & Shaw, 2002). The standard deviations (*SD*) were reported for each of the three seasonal testing periods for the third grade national norms. The fall or beginning of the year (BOY), the mean was 79.89 with a *SD* of 35.91 with a. The winter evaluation period, the mean was 96.30 and *SD* was 37.09. The last screening of the year in the spring, or end of year (EOY) mean score was 107.20 with a *SD* of 38.26. For the current study, the BOY and EOY scores were used to address overall yearly student academic growth in reading. The midterm scores were not utilized in the current study.

The technical characteristics of the *DORF* were confirmed by numerous studies using Curriculum-based Measurement (CBM) as the comparison. Reliability studies using elementary students reflected test - re-test coefficients of .92 to .97. The alternate-forms reliability of the various passages used for the fluency probes for the same levels varied from .89 to .94. The criterion-related studies conducted reflected coefficient ranges of .52 to .91.

The *DORF* was administered by an individual who was trained to administer the measure. The administration of the tests was not conducted by the students' teacher. Another teacher was appointed for this task. The PM scores were based on the average score of three one-minute probes that were read by the student while the proctor followed along and marked incorrect or skipped words read. The *DORF* raw scores were used as seasonal benchmark goals for grade-level assessing and provided PM between the seasonal testing. The *DORF* results were reported as raw scores that were based on an extensive United States cross section sample and were organized by grade-level expectations. The raw scores of the students were then compared to cut-off results that were based on oral reading performance curves calculated by the DIBELS researchers that were derived from national norms.

Performance levels of the third-grade *DORF* were based on cut-off scores that were calculated by the authors of the assessment instrument using the Receiver Operating Characteristics curve. Students were given descriptive categories of At-Risk, Some Risk, and Low Risk based on the earned score. An example of a *DORF* score of less than 53 at the beginning of the third-grade school year was described as At Risk, indicating a student who was presenting with deficits that would reflect the need for intensive

interventions. Scores between 53 and 77 would suggest Some Risk, with instructional recommendation of strategic interventions. The nationally-normed scores that were at or above 77 words per minute at the beginning of the year reflected grade-level expectations for third-grade students. These scores reflected national norms based on a sample of 9,662 third grade students in 78 school districts across the nation. These scores reflected the nationally-normed third grade cut-off scores for the *DORF*.

The Performance Series, 9th Edition Reading (*PSR*) measure was a computer-assisted administration that was provided through a website host. The Cronbach's Alpha Coefficient reflected a mean Standard Error of Measurement of .29 and a reliability coefficient of .91. A 12-week window was recommended between testing intervals. Administration time was approximately one hour for the student to complete a testing session that included a math section. The current study only used the reading scale score, which is a composite score based on the level of the test, fluency, and vocabulary skills. The tests were computer adaptive based on the student responses to the intervention content. The content was increased or decreased in complexity in order to meet the student at an ipsative level. The scaled score provided a reliable estimation of the student's reading skill level based on the statistical Rasch model (Bond & Fox, 2012). The Rasch model scores are independent of the student's grade level, and allowed for the assessment of student reading skills relative to the achieved difficulty level of the passage.

The Rasch model utilized scaled scores. These scale scores were based on grade-level norms of students across the country and ranged from 1300 to 3700. Third grade students' reading scaled score at the 50th percentile in September was 2317, in January or

middle of the year (MOY) was 2396, in May was 2467 based on national norms with a SEM of 34. For the current study, the MOY scores were not used intentionally. The MOY scores were not pertinent to the scope of the current study; only *beginning of the year* and *end of the year* were necessary in order to identify performance gains over the entire school year. An expected trajectory of reading performance scale score gain based on the national norms for a third-grade students who scored at the 50th percentile was 212 points between the fall and spring testing period. The standard score difference reflected the students' overall yearly progress. The reading scale score was used as a means to monitor the progress of students over time. The scale scores were derived from comprehension, fluency, and vocabulary development subscales that the student attained on the computer-assisted test.

When the student opened the first page of the reading assessment of the *PSR*, the student was exposed to a written passage on the computer monitor that was based on the student's current grade level and time of year the test was administered. When the student completed reading a passage, a set of multiple-choice questions was presented. The content of the questions was comprised of reading comprehension, vocabulary skills, and fluency rate provided in a multiple-choice format. The questions increased or decreased in complexity based on the student's responses. Once the test questions were completed or the child failed several of the questions, the test software computed scale scores based on national norms. The scaled scores were based on a progressive grade-level score that increased or decreased based on student's correct responses to the questions. The scores for the subtests on the *PSR* were based on projections of expected grade-related academic performance known as the Rasch Model (Bond & Fox, 2012).

The initial process of the identification of student need was based on the benchmark testing results and consisted of several steps. Each step provided the necessary building blocks that were analyzed in order to convert the data into meaningful and relevant information. The first step identified the third-grade population of the school district's elementary schools. The sample data of the *DORF* raw scores and the *PSR* scale scores provided two separate forms of evaluation data. The *DORF* data was based on raw scores. The *PSR* scores were based on normed-based scale scores. The means of these scores were calculated for each seasonal benchmark-testing period. The mean scores of each grade level based on the *beginning of the year* and *end of the year* seasonal testing periods were computed for the current study in order to obtain yearly reading progress information. If interventions were provided to the student, a determination of what level of reading support was needed.

Directed interventions were provided and were specifically geared to the student's reading development. The interventions provided the basis of remediating the child's reading needs as required by the regulations stated in IDEA 2004. This law clearly defined the need to provide a means to identify the children with reading deficits, provide researched-based interventions, and monitor the academic progress of the students. If the child did not respond to the initial interventions, the intensity of the interventions was increased.

The next phase of the statistical computations required the standardization of the nationally-normed student responses for analysis that converted the raw scores from the *DORF* scores and the scale scores of the *PSR* to *Z*-scores. The *Z*-scores were used as a means to compare the scores of the measures to each other. Local norming was then

developed to better assess the children in the study population. Local norms were used and provided a diminished bias on a local population for the identification of potential reading problems instead of using national norms (Stewart & Silberglitt, 2008).

The analysis of the *DORF* scores was conducted by the authors of the *DIBELS* tests. These instruments were nationally-normed based on the analysis of the Receiver Operating Characteristics (ROC) statistic. The ROC curve was used to identify cut-off scores that supported decision making based on the student's PM benchmark scores to assess progress or lack of progress in reading fluency performance. These nationally-normed scores provided cut-off scores that limited false detections and supported the identification of students with reading deficits (Briggs, 2011; University of Oregon, Center on Teaching and Learning, 2012). The scores were labeled into categories or levels to establish threshold values for students with various levels of risk or potential learning disabilities (Good et al., 2004). Although the ROC scores supported the identification of cut-off values that aided decisions by distinguishing optimal cut-off scores for a given sample, the scores were nationally normed and did not take in account communities with a higher proportion of culturally diverse students. The norming was based on a cross section of the national population.

The Illinois State Board of Education (2012b) survey data for the school year 2010-11 identified 4.6% of the 2,087,628 students, or 96,031of the students with learning disabilities in reading. The same year, evaluations utilizing national norms with the student population of the school district researched in the current study presented with higher levels of eligibilities for learning disabilities 6.4%. This higher identification rate of SLD prompted the development of local norms with the cut-off value based on the 10^{th}

percentile, or *Z*-score -1.27, as the method to identify the statistically significant discrepancy of the students through the utilization of the universal screening measures to assess for potential learning disabilities. Although this method was useful for the identification of a severe discrepancy, it lacked the information necessary to classify the students who were not responding to the MTSS.

The assessment of students' academic performance, was reflective of the PM scores that were achieved over time. Both the *DORF* and *PSR* scores yielded the data that identified the level of progress the child made. Neither one of the scales used slope scores to assess the student's level of performance over time. Statistical gradient was used for the computation the student's rate of improvement. It was based on the Tukey method of calculation for slope. The Tukey slope method was utilized as a means to identify the progress made by the student over time. This calculation was made by subtracting the *end of the year* score from the *beginning of the year* of both the *DORF* and the *PSR* scores and then dividing the numerator by the number of days elapsed between the testing periods, which for this study was 170 elapsed school days and were not chronological days. The difference between school days and chronological days was based on actual days that school was in session versus consecutive days of the year. The method resulted with a coefficient score that was compared to the general mean of the population over time. Scores that reflected a lower coefficient than the mean, did not reflect an upward trend, and indicated a poor response to the interventions were indicative of SLD.

The utilization of the Chi-Squared statistic was necessary to determine the frequencies of correct diagnoses of specific learning disabilities when reviewing the local scores with the actual students formally diagnosed with reading learning disabilities.

Further evaluation needed to incorporate the integration of both the performance standard scores and the slope data. More specifically, the local norms would perform as the source of this evaluative information that was to identify the frequencies of actual diagnoses.

The scores of the *PSR* and *DORF* were converted to standard scores using *Z*-scores. Descriptive statistics included the sample, mean, and standard deviation. Cut-off scores were determined using the 10th percentile, or a *Z*-score of -1.27. The combined classification of SLD using the *PSR* and *DORF* measures was calculated using the contingency coefficient statistic. The *end of the year* data was correlated with the performance and slope scores based on local norms. The examination of the correlation between the discrepancy of the measure scores and slope scores was conducted to determine the presence and strength of the relationship between the identification of reading–based learning disabilities verified by local norms.

Summary

Universal Screening evaluations used to assess the progress of children in schools provided a meaningful and efficient way to ensure children were not left behind in the educational process. The difficulty lay with the question of how the data was utilized to monitor student academic performance in the general education curriculum. Standardizing the scores of the screening or PM data, identifying the level of statistical significance required for identification of academic needs based on local norms, and establishing standardized trajectories of PM scores would support the eligibility decision making process for SLD. Developing a standardized system to interpret the data using *Z*-scores was fundamental for the comparison of the scores relative to different PM measures. The use of cut-off scores supported the evaluation and classification process of

students with potential learning disabilities. Cut-off scores in conjunction with slope scores were used to assess students who were not responding to the interventions provided through MTSS. The current study attempted to utilize a method of using standardized locally-normed reading PM measures that were used with third-graders in a school district with high levels of minority students living in a lower socioeconomic community to develop consistent decision that limited disproportionality eligibility for SLD (Illinois State Board of Education, 2009).

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The assessment of learning disabilities (LD) has had a long history of ambiguity (Sandoval, 1993). Often, it was very difficult to understand why some children fell behind academically (Wasnek & Vaughn, 2007). Furthermore, Reynolds and Shaywitz (2009) asserted that a range of conditions and reasons that prevented children from learning fell in the category of a Type I error that identified a specific learning disability when the student did not meet the eligibility. In light of this statistical error, the evaluation proved problematic with these students who were impacted by social circumstances, culture, or language differences (Utley et al., 2011). These evaluation problems ranged from the environmental forces that acted upon children such as personal, familial, and social problems to genetic phenomena that have yet to be understood (Appelbaum, 2009; Lee & Stewart, 2013; Rosenberg, Bart, & Ratzon, 2013).

Based upon the childhood and adolescent learning disorders addressed in the diagnostic manual from the American Psychiatric Association (2000) and the data collected by the Centers for Disease Control and Prevention (2011), students with specific learning disabilities comprised a range of 5% to 15% of the school-age children across the United States. The most prevalent form of identified LD involved reading disorders. Of children identified with a LD, a reading disorder was noted to impact four

out of five children alone or in combination with other learning disabilities (American Psychiatric Association, 2013; Illinois State Board of Education, 2012b; Reschly & Ysseldyke, 1995; Tran, et al., 2010). Many of these children who were identified with LD came from environmental situations and conditions that resulted from poor student response to the curriculum.

The increased frequency of the identification of LD eligibilities reported to state boards of education across the country prompted legislative changes that impacted the assessment process of LD (United States Department of Education, 2013; Jenkins, Schiller, Blackorby, Thayer, & Tilly, 2013, Zirkel, 2013). Although previous research had attempted to address the issues related to the evaluation of LD and how the PM data was interpreted by school professionals, assessment practices have presented as inconsistent and ambiguous (Berkeley, Bender, Peaster, & Saunders, 2009; Fuchs et al., 2007; Scanlon, 2013). The statistical relevance of discrepant PM scores supported the criteria for LD eligibility. This relevance was due to the evaluation of students over a duration of time and evidence of not responding to interventions based on PM scores. However, a systematic process that determined a reliable and valid performance discrepancy threshold scores was required for the students with potential LD (Hixson et al., 2008; Hosp et al., 2011).

In the past, the practice of LD identification was a one-size-fits-all assessment using the ability-achievement discrepancy theory model. The standardized practices presented with theoretical shortcomings that often resulted in inaccurate identification of students in diverse minority populations that was a reflection of Type I statistical errors (Callinan, Cunningham, & Theiler, 2013; Harry & Klinger, 2006). Type I Errors were the

results of diagnosing a child with LD when he or she was not. The impact of socioeconomic influence on socially and culturally diverse schoolchildren was a contributing factor that negatively impacted academic performance (Utley, 2011; Utley et al., 2011).

Often, students from diverse cultural and lower socioeconomic communities presented with academic limitations that were improperly diagnosed as learning disabilities (Klinger & Edwards, 2006). These students were often identified with a disability for the sole purpose of increased academic support (Harry & Klinger, 2007). In these cases, school districts were often directed by their state board of education to remedy the disproportionality of minority students who were identified as having specific learning disabilities (Illinois State Board of Education, 2012b). Part of this intervention included the restructuring of the methods used to identify SLD. No longer could the school psychologists use one method to diagnose learning disabilities. United States Department of Education (2013) Individuals with Disabilities Education Improvement Act of 2004 (IDEA) legislation confronted the identification process of SLD evaluation, which addressed the need for a systems approach to learning disabilities assessment practices.

Learning Disability Evaluation Process

Since the initial drafting of *Public Law 94-142,* the definitions of the various eligibilities have evolved. One of the areas of substantial changes to IDEA included the process of evaluating Learning Disabilities (LD). The definition of LD presented with a broad spectrum of deficits that impacted neurological processing, especially related to the manner with which an individual learned to read, developed math calculation and

reasoning skills, and used receptive and expressive language (Floyd, Keith, Tuab, & McGrew, 2007; Hale, et al., 2010; United States Department of Education, 2013). A learning disability was described as a disorder in one or more of the basic psychological processes involved in the comprehension or use of language, spoken, or written (Zirkel, 2010). These processes were manifested as the imperfect ability to listen, think, speak, read, write, spell, or calculate mathematical operations (Tannock, 2013). Other conditions included perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia (United States Department of Education, 2013).

In order for a child to qualify for a learning disability eligibility, specific criteria must be met, subsequent to Individuals with Disabilities Improvement Act. The basis of an LD diagnosis cannot be solely determined by the one form of evaluation (Cavendish, 2013; Epps, Ysseldyke, & McGue, 1984; United States Department of Education, 2013). This discrepancy model was studied by various researchers who found it lacked external validity and was therefore avoided as a means for the determination of learning disabilities (Dombrowski et al., 2004; Kuntz & Eulass, 2009; O'Connor & Klingner, 2010; Fletcher, Coulter, Reschly, D. J., & Vaughn, 2004; Francis et al., 2005). The student with a possible LD diagnosis reflected a profile of the student who failed to meet grade-level academic progress despite general education interventions through differentiation of instruction. Intervention failure, based on the three-tier system, was the final requirement in the evaluation process for SLD (Gersten et al., 2009; Mellard et al., 2010). General education curriculum standards and criteria had to be established before LD was considered (Brown-Childsey & Steege, 2010).

The children who necessitate interventions required increased periodic testing to identify academic progress (Fuchs, & Fuchs, 2007; Hixson et al., 2008; Mellard et al., 2010). Monitoring the child's proximity to the average student progress was essential for the assessment of the potential LD profile (Therrien & Hughs, 2008; Tichà, Espin, & Wayman, 2009). The ability to assess the child accurately through the utilization of a variety of assessment scores was required (O'Connor & Klingner, 2010). This process of evaluation required careful attention to ensure the consistency of a learning disability profile. The PM scales, which have been nationally normed, validated, and found to be reliable, provided the universal screening instruments for school administrators and teachers to follow student progress over time (Hagermoser-Sanetti, Gritter, & Dobey, 2011). These measures were used as specific means to identify targeted areas of need for the students. The main focus of the measures was to identify which children may need further interventions or have the potential for a disability covered under IDEA legislation (Deno et al., 2009; Ysseldyke, Burns, Scholin, & Parker, 2010). Each of these measures provided scores that required tables and charts to identify and classify the child's current level of academic performance (Bocian, Beebe, MacMillan, & Gresham, 1999; Compton et al., 2010; Fletcher et al., 2011).

Z-scores were utilized to standardize the multiple related measures of the PM data in order to compare results. Standardizing the various scores used for PM was important because many of the scores were reported in different formats. *Z*-scores have a mean of 0.0 and standard deviation of 1.0. Each of the scores were converted to *Z*-scores reflected the provided a method to compare the students' results on the measures directly. This scoring method eliminated the various scores from the benchmark tests. The results were

readily compared based on standardized scores. The importance of standardized data, when local norms were utilized, allowed for the evaluators to compare and interpret the PM data reliably (Christ & Silberglitt, 2007; Hixson et al., 2008). Comparison of the tests administered based on the standard scores helped minimize the misidentification of children with potential learning disabilities (Hale et al., 2010; Johnson et al., 2009).

When the team of interventionists determine critical score deficits in children using a procedure-based policy, an increased rate of consistent decision-making was evidenced (Ball & Christ, 2012; Compton et al., 2010; Shapiro et al., 2011). The determination of statistically significant score discrepancies based on the locally-normed PM measures between the student who was in the process of being evaluated for potential LD and the average students' scores were essential to address the diagnosis of LD (Yeo et al., 2011).

Kavale (2005) remarked about the importance of identification profiles for LD, which would increase the efficacy of decisions by school personnel for students who were not responding to interventions. The use of the PM results with slope scores provided additional information for the profile of discrepant scores compared to the student's peers. However, without statistical clarification of the potential profiles of children with learning disabilities, no consistent team decisions were derived from the data (Kavale & Spalding, 2008; Kim, Petscher, Schatschneider, & Foorman, 2010; Shinn, 2001). The measures used for PM within a given local school setting were based on various scales that assisted the decision-making teams to interpret a student's relative academic progress (Compton, Fuchs, Fuchs, & Bryant, 2006).

Interval assessment of progress was one of the primary formats used to measure how children responded to the reading curriculum (Ardoin, 2006; Ardoin & Christ, 2009). Various research-based instruments or measures that assessed progress included, but were not limited to, CBM and various proprietary assessments were designed to assess reading skills (Christ & Coolong-Chaffin, 2007). Many of the tests reported a composite score that included reading fluency skills, vocabulary development, and comprehension skills. Although the data helped to guide the intervention decisions of educators, the data also provided a source of information, which supported individual learning needs (Christ, Zopluoglu, Long, & Monaghen, 2012).

The majority of the psychometric and achievement measures that the standard scores were based on national norms were recognized by numerous researchers to not fairly assess the needs of the children, in particular, socioeconomic and culturally diverse populations (Pearce & Gayle, 2009; Prasse & Reschly, 1986; Utley et al., 2011). Local norming provided the response to the problems associated with the data needed for curriculum driven interventions for the school district or school and individual student level (Stewart & Silberglitt, 2008). The problems that assessors struggled with were the differences of how the scores were presented. Each measure was reported in isolation even though the scores provided similar information about the student's achievement. A combination of standardized scores in conjunction with slope data of the individual's progress was indicated. The utilization of the universal screening scores to compare and contrast provided the essential eligibility assessment information for LD profiles based on IDEA 2004 legislation.

The next step for LD eligibility process required a comprehensive evaluation. The educational diagnostician must utilize an assortment of evaluation measures. Several forms of assessment information are required in combination, including intervention response, PM, interviews with teachers, as well as other forms of assessment information, in order to diagnosis a learning disorder (Kaminski & Good, 1996; Scanlon, 2013; United States Department of Education, 2013). The measures included CBM, universal testing, reading skills assessments, and other similar evaluation measures. These measures also included intelligence quotient (IQ) and norm-referenced achievement instruments, which have historically been utilized for the evaluation of learning disabilities.

Exclusionary criteria had to be ruled out before the student was found eligible for a diagnosis of LD. Exclusionary criteria reflected a list of medical and environmental factors that identified the underlying reason for the learning problems, and were most likely the primary cause for the lack of progress or learning. For example, if the student had another disability that impacted vision, hearing, or motor skills; intellectual disability; emotional disorder; cultural factors; environmental and/or economic disadvantage; or English proficiency limitations, then he or she may not have qualified for an LD diagnosis (United States Department of Education, 2013).

The Misidentification of Learning Disabilities

The limitations of the discrepancy model included the misidentification, or disproportionality of LD, especially with regard to minorities (Epps et al., 1984; Harry & Klinger, 2007; Sternberg, 1985; Sullivan & Artiles, 2011). Type I and II errors were commonplace when assessing children with various environmental and socioeconomic limitations (Steubing et al., 2002; Van Der Heyden, 2011). These statistical errors were

mainly based on the identification of a learning disability when there really was not, or Type I error. Various modifications that ranged from the utilization of discrepant standard deviations for the chosen instruments to regression formulas developed from the standardized scores from the selected measures were explored (Francis et al., 2005). The research to address the difficulties that were inherent with the use of the ability-achievement model began in earnest. The trend of misidentification of learning disabilities with children from diverse environmental, cultural, and socioeconomic circumstances was also examined (National Center for Response to Intervention, 2009; Utley et al., 2011).

The LD evaluation process has evolved over the last 40 years. Different approaches for assessing learning disabilities through the use various models and constructs have been utilized and extensively researched (Ahearn, 2009; Ysseldyke et al., 1983). Prior to the reauthorization of the public law that included the Individuals with Disabilities Educational Act of 2004, researchers hypothesized that the IQ was not inconsistent with an individual's educational achievement (Aaron et al., 2008; Callinan et al., 2013; Fletcher et al., 2011). However, when a difference between the measures where the achievement standard score was statistically significantly below the IQ score, a learning disability was evident. This form of evaluation was known as the ability-achievement discrepancy model. Simply stated, when the individual's IQ standard score fell within the average range and achievement scores were at least a standard deviation or lower, a learning disability was presumed (Stoodley, Ray, Jack, & Stein, 2008).

The ability-achievement discrepancy model was used exclusively despite growing concerns regarding validity and reliability of the model (Reynolds & Shaywitz, 2009).

Researchers began to evaluate the ability-achievement discrepancy model and found indications of problematic limitations (Ysseldyke & Thurlow, 1984). These problems focused on the Type I and II errors that were often related to the misidentification of minority students living in urban areas of the country (Prasse & Reschly, 1986).

One of the primary concerns with the ability-achievement discrepancy model was the concern with intelligence measures underestimating intellectual functioning of minorities. Sternberg (1985) developed three sub-theories known as the Triarchic Theory of Intelligence that presented underestimation of IQ scores based on mainstream intellectual functioning instruments. This underestimation of IQ led to concerns of the misdiagnosis of intellectual disabilities in minority populations. This theory of intelligence was an early attempt to develop a measure to assess intellectual skills with a diminished bias towards norm sets based on the middle class. Sternberg postulated that an IQ assessment required the full range of constructs related to intelligence instead of the crystallized formats of the accepted IQ assessments of the day. His research was an effort to limit the misidentification of minorities with intellectual disabilities (Sternberg, 1983).

The intellectual functioning research of diverse populations in impoverished inner cities led to the identification of certain psycho-cultural differences that negatively impacted the educational process of many children in lower socioeconomic communities (Prasse & Reschly, 1986; Sternberg, 1983). The leading principle addressed in this research concentrated on how intellectual performance was socio-culturally relative to the individual’s experiences. The link between the direct training experiences and the development of strategies and coping styles of the individual provided them with the

capability to address novel situations and activities (Salvia & Ysseldyke, 1998; Sternberg, 1985).

The experiences of children from these lower socioeconomic communities were different than the middle-class children that the IQ testing was primarily normed (Prasse & Reschly, 1986). Therefore, individuals with diverse cultural experiences tended to achieve lower IQ scores than the middle-class norm sample. The minority students with socio-cultural diversities based on environmental factors who were tested using the ability-achievement assessment methods were more often misidentified with LD or other cognitive/intellectual deficits (Forsyth, Colver, Alvanides, Wooly, & Lowe, 2007).

Through the 1980s, research about the bias of the ability-achievement method came under scrutiny, which led to researchers refuting this methodology for the identification of learning disabilities (Ysseldyke & Thurlow, 1984). Within this time frame, federal court judges in California were overturning decisions made by public schools regarding the use of intelligence scales with minority children (Prasse & Reschly, 1986; Reschly, Kicklighter, & McKee, 1988). According to the ruling of the federal judges who heard the cases, intelligence scales did not reliably measure global intelligence of minority students. This action gave rise to an important question of validity and reliability with the use of the ability-achievement model when utilized with minority children.

Another method of valid and reliable evaluation was required in order to eliminate the biased nature of the national norms inherent to the ability-achievement model (Christianson & Ysseldyke, 1989; Speece, 2008). School psychologists began to look elsewhere for other evaluation instruments. A comprehensive form of evaluation that

included school-based academic interventions was needed to assess the child over time that would provide data to support educational decision making, rather than a one-time psycho-educational battery evaluation method (Tichà et al., 2009).

Curriculum-Based Measurement

CBM was developed in the late 1970s to the early 80s. The primary use of CBM evolved from the research about LD (Ardoin & Christ, 2009; Christ et al., 2010; Deno, 1985; Hale et al., 2010; Swets, Dawes, & Monahan, 2000). This form of evaluation identified students who were struggling academically within the school system through the utilization of universal screening practices (Fuchs & Fuchs, 2005). The development of local norms for a particular curriculum was based on the utilization of probes and benchmark testing, and was used to monitor student progress over time rather than evaluate and diagnose (Canter, 1995; Germann & Tindal, 1985; Hadedank, 1995; Marston & Magnusson, 1988; Shinn, 1988).

Another essential contribution that CBM provided was that children were not as susceptible to the *wait for failure* method of the ability-achievement discrepancy model (Fletcher, Coulter, Reschly, & Vaughn, 2004). Academic subject-specific tests known as probes took minutes to administer and were developed to monitor the progress of a child at specific grade levels. Interventions were provided to support the student instead a waiting until the discrepancy was large enough to consider them for identification of a learning disability. Benchmark scores provided data to assess and compare student progress throughout school year in order to catch the struggling students earlier instead of later (Christ et al., 2013; Hintze & Silberglitt, 2005).

The practice of interval assessment throughout the academic year provided school personnel with a running record of evaluation probes that were presented in a graph and report format. More specifically, universal screening evaluations were developed to assess reading, writing, or math performance skills of the students (Germann & Tindal, 1985; Tichà et al., 2009). Scores based on the probes were graphically represented to assess whether or not the students were responding to the instructional practices of the teacher or mastering the curriculum content (Good et al., 2001; Marston & Tindal, 1995; Shinn, 1989; Shinn & Marston, 1985).

Taking the CBM practice one step further, school personnel would provide interventions directed at the specific academic needs of the child, based on the school curriculum (Deno, 1985; Hosp, Hosp, & Howell, 2007). The measurement of progress, within the curriculum, was founded on benchmark data. This benchmark data limited ambiguity and provided reliable scores of the ordinal information based on student academic performance. To ensure simplicity and consistency, models were developed in order to understand or interpret the findings of the CBM. These education-based models were inherently more useful through an assessment over a period of time rather than the one-shot, theoretically-based, norm-referenced intellectual or achievement evaluations (Salvia & Ysseldyke, 1998; Sibley, Biwer, & Hesch, 2001). The assessment data reflected objective progress based on criteria developed through the actual curriculum (Chun & Witt, 2008; Deno et al., 2001; Jenkins & Terjeson, 2011; Messick, 1995; No Child Left Behind Act, 2001; Shinn, 2001; Ysseldyke et al., 2010).

The application of CBM to assess and support student learning was supported by several instructional models. These models addressed factors that contributed to

academic success in the school environments. Carroll (1963) developed one of the first models that provided a foundation for many of the CBM models that have been developed to date. The focus of Carroll's Model addressed how children learn through a process. This process was divided between the actual time the child was learning and by the quality of the instruction that the child required to understand the curriculum material. The student's ability to learn and the quality of the instruction were the primary indicators of the degree, which children would successfully learn. This model led to research that began to encompass environmental conditions and factors that influence learning.

Another model developed by Algozzine and Ysseldyke (1992) focused on the characteristics of student outcomes that related to the curriculum of the school, the conditions within the school district and school, and family environmental factors. This model focused on clarifying and presenting what was needed to ensure the success of students in the learning environment. The characteristics and conditions that included instructional planning, student management, instructional delivery, and the evaluation of students were identified as the principle strategies to support children with their learning (Algozzine, Ysseldyke, & Elliott, 1997). This model was one of the precursors to the changes that were translated into legislation for IDEA 2004.

School psychologists began to adopt the RTI method, based on the inferential evidence of CBM research. This intervention method allowed for support of the teaching personnel in the classroom environment, rather than on the sole evaluation by the school psychologists (Fagan, 1995; Reschly & Ysseldyke, 1995). The efficacy-directed academic interventions allowed for the remediation of skill deficits and the identification

of specific learning disabilities, which further enhanced academic supports for the students (Marston & Tindal, 1995; Reynolds, 1992; Ysseldyke et al., 2010).

Response to intervention provided the framework for universal monitoring of students within a multi-tiered system of support (MTSS) (Bamonto-Graney, Martinez, Missall, & Aricak, 2010; Mellard, McKnight, & Woods, 2009). The importance of making data-driven decisions for children who were struggling academically was hallmarked by the use of RTI or MTSS procedures. Schools began the utilization of various evaluation tools for the interval assessment of student progress. The importance of making valid and reliable estimates of student academic performance was paramount (Bramlett et al., 2010; Hosp et al., 2011; Jenkins, Graff, & Miglioretti, 2009).

The procedure of universal screening for reading supported the identification of students who were not responding to the general education instruction (Brown-Childsey & Steege, 2010; Glover & Albers, 2007). Interventions were necessary to ensure that students who were not responding to the general education instruction were not doing so because of a lack of instruction and/or environmental, cultural, or motivational concerns (Callinan et al., 2013; Fuchs & Fuchs, 2007; United States Department of Education, 2013; No Child Left Behind Act, 2001). Interventions were provided that addressed the individual instructional needs based on the student’s response to the curriculum before LD was determined (Batsche et al., 2008; Kavale, 2005; Kavale & Spaulding, 2008; Mellard & Johnson, 2007; Parker et al., 2012).

The determination of whether or not a child had a specific learning disability required an established infrastructure and protocol where school districts or agencies identified whether a child responded to research-based interventions (Brown-Childsey &

Steege, 2010; Fuchs & Fuchs, 2005; Hagan-Burke & Jefferson, 2002; United States Department of Education, 2013). The intervention process, determined by CBM and other relevant measures, would support student progress or lack of progress over time (Aaron et al., 2008; Good et al., 2001; Hagermoser-Sanetti et al., 2011; Jenkins et al, 2009; Shinn, 2007).

Assessment Changes for Learning Disabilities

A substantial problem, however, remained concerning the assessment of LD (Christianson & Ysseldyke, 1989; Deno et al., 2001; Francis et al., 2005). The problem was that many of these children were struggling students who may have had difficult environmental circumstances, new language acquisition, and limitations of the exposure to academic material. This limitation of academic curriculum and cultural exposure often led to the misidentification of learning disabilities within the school systems. Often, these children would respond poorly to the norm-referenced testing that was the primary source of evaluation prior to the reauthorization of IDEA 2004.

The utilization of the ability-achievement discrepancy model was the root to the misidentification of LD in many minority students that presented with bilingual and/or lower socioeconomic conditions (Harry & Klinger, 2007; Utley, 2011). School psychologists continued to diagnose LD in the school environment with a one-time evaluation. Clinical psychologists in the community would evaluate a student with the same model and provide parents with the report indicating eligibility for special education services due to LD. The problem was that many of these children were simply struggling students misdiagnosed with a learning disability who mainly struggled with academic motivation (Christianson & Ysseldyke, 1989; Polcyn, Levine-Donnnerstein,

Perfect, & Obrzut, 2014; Prasse & Reschly, 1986; Ysseldyke et al., 2010). The discrepancy evaluation format often was commonly referred to as the wait for failure model (Fuchs & Fuchs, 2006). A change in the process of assessment for LD was warranted.

In 2004, IDEA was reauthorized by Congress to include changes in the process of assessment and diagnosis of LD. The LD eligibility continued to be defined in an ambiguous manner due to the problems associated with the operational definition and the other factors that had a negative impact on learning (Fuchs et al., 2007; Speece, 2008). One of the most important changes to this landmark legislation, however, was to revamp the LD eligibility criteria and to eliminate the dependence on the ability-achievement discrepancy model for the diagnosis of LD (Dombrowski et al., 2006). Exclusionary criteria were also added to the legislation as a means to address situational, medical, and environmental considerations. The use of IQ or ability as an indication of learning potential was considered a one-shot model that did not attempt to remediate through interventions, but rather it was a wait for failure method (Brown-Childsey & Steege, 2010).

In order to adhere to the legislative criteria for LD, the school must use state approved grade-level standards. Although the LD criteria were identified in the legislation, the school districts were granted the authority to develop the means of identifying LD for the particular population (Mellard, McKnight, & Woods, 2009; United States Department of Education, 2013; Wedl, 2005; Zirkel, 2010). Students were not identified with learning disabilities if the problems evidenced were directly a result of hearing, visual, motor-related disabilities, intellectual disabilities, emotional disabilities,

or a result of environmental, cultural, linguistic, or socioeconomic disadvantage (United States Department of Education, 2013). In order to rule out these conditions, MTSS practices considered these factors and accommodated the students by providing the appropriate instruction to ensure success. It was only when the student did not respond adequately to these interventions that LD was considered as a factor contributing to the child's learning problems.

Utley et al., (2011) investigated culturally sensitive practices in order to address the needs of children with cultural and linguistic diversity who were struggling academically. The use of psychometric instruments did not fully illuminate the students' potential who presented with these social and cultural diversities. Therefore, the nationally-normed assessments were not adequate to evaluate culturally diverse children with LD (Sullivan, & Artiles, 2011). The utilization of local norms using specific assessments designed to assess reading, math, and writing was needed to compare student progress (Marston, & Magnusson, 1988; Stolar et al., 2008).

The Case for Research-Based Instruction

The general education curriculum required research-based instructional content provided by highly qualified certified teachers (Kuntz & Eulass, 2009; No Child Left Behind Act, 2001). The prerequisites of the teacher qualifications were the fundamental step in the process to ensure quality instruction (Darling-Hammond, 2000). According to the NCLB definition, a highly qualified teacher had met the state standards of professional development in a given core academic subject area (Karelitz et al., 2011). The criterion of highly qualified teacher certification was mandated to help ensure that

quality instruction was being provided in the public schools (Chin & Wong, 2013; No Child Left Behind Act, 2001).

By setting high standards for teacher instructional effectiveness, the public school systems ensured that a research-based curriculum was provided to the children with efficacy (Stecker et al., 2008). This investment in the professional development of teachers was a major investment by the government through No Child Left Behind Act of 2001 legislation to improve the educational performance of students, especially in the area of reading proficiencies. There was still a concern regarding how to provide the universal means of monitoring the progress of student reading skills development.

The RTI documentation required for LD eligibility determination contained a statement of the problem and objective data to support the assessment findings (Ahearn, 2009; Jenkins et al., 2013). The data provided in the report must also reflect whether or not the child had responded to previous interventions (Aaron et al., 2008; Roehrig, Duggar, Moats, Glover, & Mincey, 2008; Shinn, 2007). Interventions in the school system related to RTI involved three tiers of support levels. The first tier was based on the general education instructional environment. When the child did not respond to the first tier of universal supports, then the student was referred for Tier-2 level support interventions. Tier-2 was provided in a small group setting of four to eight children and was conducted by a certified teacher who specialized in that given area of academic intervention. If the child did not adequately respond to the second level of interventions, the student was referred to Tier-3 intensive interventions. This third level of interventions consisted of one to four students working on specific individual deficit areas with

certified personnel as interventionists (Appelbaum, 2009; Brown-Childsey & Steege, 2010; Swanson, Solis, Ciullo, & McKenna, 2012).

The identification of students who needed academic intervention was done through various screening methods. The use of universal screening methods administered to all of the students throughout the school year allowed for a continuing source of data with which to monitor student progress (Hosp et al., 2011; Johnson et al., 2009; Shinn, 2007). The utilization of local norms took into effect the local population diversity. Environmental limitations and cultural impact of the students from lower socioeconomic communities often led to the misidentification of LD in these children (Harry & Klinger, 2006). In order to limit misidentification of LD of these students, the evaluation process included using locally-normed benchmark interval data (Stewart, & Silberglitt, 2008; Stollar et al., 2008).

Benchmark testing provided interval data based on a specific time period during the regular school year. Seasonal universal screening evaluations were conducted in the beginning or fall of the academic year, at the mid-point of the academic year, and the end of the academic year. These benchmarks aligned with regular education-expected reading growth patterns for the academic year. Midpoints between the benchmark or seasonal testing periods, PM using the *DORF* and *PSR* helped to establish the pattern of academic growth for the student, based on the response of the child to reading skills instruction (Ardoin & Christ, 2009; Christ et al., 2010; Good et al., 2004).

When academic deficiencies were identified through the benchmark testing results, interventions were provided to help support and improve the students who were struggling to achieve grade-level reading expectations (Jenkins et al, 2009; Scott &

Weishaar, 2003). The interventions provided the basis for the remediation of student reading disparities when compared to grade level and time of year expectations as required by IDEA 2004. This law clearly defined the need to provide a means for identification of the children with reading deficits through PM, provide research-based interventions specific to the child's academic needs, and monitor the academic progress of student interventions (Petscher, Cummings, Biancarosa, & Fien, 2013). If the child did not respond to the initial interventions, the intensity of the interventions was increased (Compton et. al., 2012).

The Utilization of Slope as an Indicator of Academic Progress

As a result of the RTI model, many schools struggled with the effective and defensible utilization of information to address high-stakes decisions that included special education-eligible disabilities. These decisions required a research-based methodology to analyze a lack of academic growth and provide a means to defend these decisions through the methodology used to compare student performance. There were several forms of slope methodologies that have shown potential, but many presented with limitations due to the inclusion of complicated calculations that were difficult to use and explain to team members and also to parents. There were two slope models that were used by researchers and educational evaluators to illustrate the progress of students, the Tukey slope model and linear regression methods. For the current study, the Tukey slope model was used based on the simplicity of the methods that presented with diminished probability of evaluator error.

The use of slope to aid in the analysis of the PM data began in the 1980s with the onset of CBM. The utility of slope measurement provided an assessment that monitored

progress based on the interval data that was collected through probes and other forms universal screening data. Slope was calculated by dividing the rise in score by the run over time. The main function of slope was to determine the relevance of change of the measured performance of the student over an extended period of time (Figure 1). The slope was compared to the progress of the mean scores of the student's particular grade level; seasonal benchmark scores were used based on the universal screening evaluations. The slope data was analyzed quickly and effectively by comparing the trend line of the student's achievement scores against the performance of the other students based on the Tukey slope method (Parker et al., 2012).

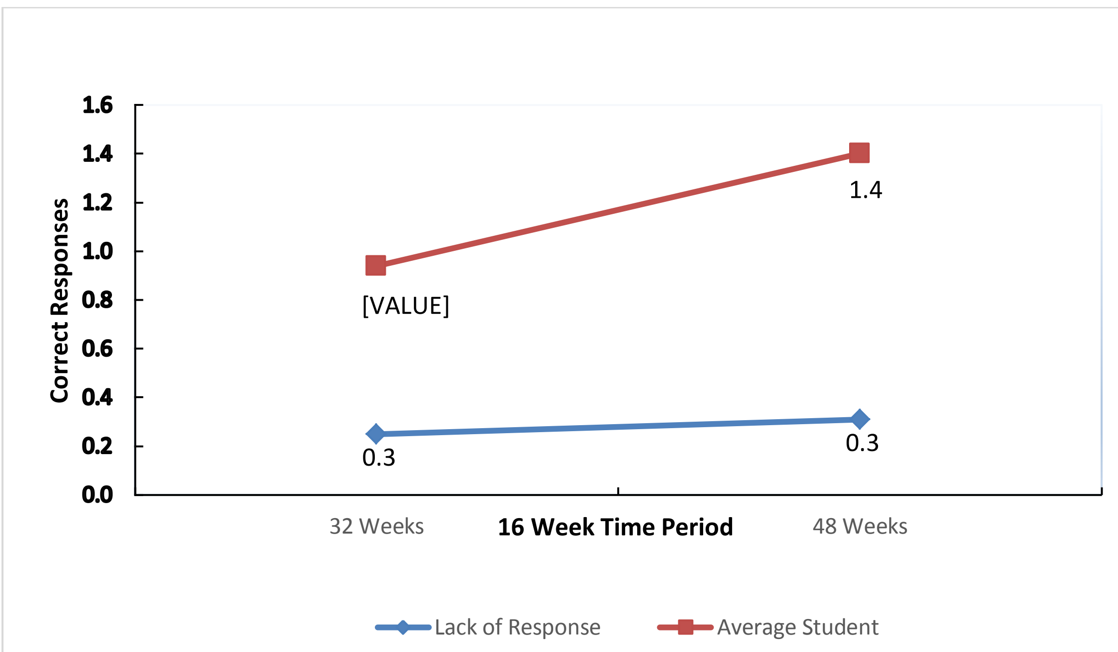

Figure 1. Sample Tukey Slope Model

The Lack of Response data reflects a limited growth over time using a Tukey slope model. Where the average student made a gain of 0.5, and the student who was not responding to interventions made a 0.0 gain based on the slope model between week 32 and 48 (sample rounded to the nearest tenth).

The most common application of slope utilized in the school systems was the Tukey Tri-Split Slope (Tukey, 1977). This application is easily hand-calculated and graphed by using a simple formula for determining slope using the change in the current PM score minus the prior or initial score divided by the current number of days minus the initial day (rise/run).

The calculation for the Tukey Slope Model is as follows:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

Where:

m is the individual slope coefficient

y_1 first score of the universal reading series

y_2 last score of the universal reading series

x_1 first assessment administration

x_2 latest assessment administration

This slope score yielded a coefficient score that supported intervention changes and potential eligibility considerations. However, the system had limitations in terms of how outliers impacted the mean scores. This issue of outliers was the reason that the MOY scores were not used in the calculation of yearly growth for the current study. The Tukey method was influenced by extreme scores based on its algorithm, which was dependent on the median. This analysis skewed scores in a negative or positive direction that presented with a few statistical limitations, but reflected the student's progress or lack thereof. The ability to assess whether or not a child responded to interventions based on a quantitative score offered a means to identify the need for further interventions or to refer the student for a special education learning disabilities evaluation.

The Significance of Percentile Score Ranges Compared to Standardized Scores

Shinn (2001) postulated that the utilization of local norms should be used in conjunction with discrepancy ratios. He suggested using cut-off scores based on percentile ranks, which offered a more recognizable source of information for most educators, rather than continued use of discrepancy ratios. The suggested score was based on 10th percentile (Christ et al., 2012; Good, Gruba, & Kaminski, 2001; Shinn, 2007). When students consistently fell below this percentile range, the student presented evidence of a potential learning disability. Further evaluation was then indicated.

Even though these percentiles scores were well below the average range, if local norms were used, there was a way to quantify how a statistically meaningful score profile was validated. If there were specified benchmark scores for the average student to show progress, students with scores that were statistically lower than other students were identified through cut-off scores. However, the practice of comparing percentile rankings has limitations when used to compare with other scores. The sample of the data tended to vary to a higher or lower degree due to the number of individuals administered the benchmark testing. The results of the benchmark testing would not reflect the same total of students who completed the prior testing, thus impacting the ranking of the individual score comparison for that particular benchmark result.

When percentages were used, the benchmark goals for the *DORF* and *PSR* only indicated that the student may need more intensive instructional support based on the results of the score that the student received. After several administrations of the testing probes, student scores should reveal and clarify how the child was responding to interventions (Deno et al., 2001; Gresham, 2009; Hasbrouck & Tindal, 2006). The

comparison of percentile scores did not lend itself to responsible psychometric practices (Salkind, 2012). The use of standardized *Z*-scores for both the *DORF* and *PSR* would provide data that was compared over time. A *Z*-score of -1.27 would equate to the 10^{th} percentile. The benefit of the *Z*-scores was that they can be compared and manipulated statistically. These scores may also help to reliably assess whether individuals demonstrated the ability to respond to a targeted intervention (Christ et al., 2012; Good et al., 2001). A question of what the valid and reliable score profile would typically look like, based on children from lower socioeconomic communities who presented with LD, needed further analysis and continued to present as relevant research questions (Bocian et al., 1999; Gresham, 2009; Utley et al., 2011; Utley, 2011; Yeo et al., 2011).

Validity and Reliability of Score Interpretation Practices

The assessment of learning necessitated an empirical and research-based PM model. Evidence was required to support the theoretical and legal rationales of inferential verification of actions that were derived from the results of the test scores (Gay, Mills, & Airasian, 2012; Glover & Diperna, 2007; Hosp et al., 2011; Jenkins & Terjeson, 2011; Messick, 1995). One of the pertinent issues regarding validity with making inferences based on several forms of assessments, focused on ensuring that appropriate decision thresholds were calculated based on the local norms (Nese, Park, Alonso, & Tindal, 2011; Swets et al., 2000). Reliable diagnostic information addressed through correlation studies assisted with the decisions based on the threshold scores derived from seasonal benchmark results (Kim et al., 2010).

The utilization of two separate proprietary universal screening tools for comparison use required similar constructs for each measurement instrument. When

using the various PM measures to assess for LD, construct validity was important (Tichà et al., 2009). The two measures used in the current study, the *DIBELS Oral Reading Fluency* and the *Performance Series Reading* scales, addressed the need for multiple assessments to support verification of the assessed student reading progress.

The scores for the *DORF* reflected a strong relationship between reading fluency and comprehension with early elementary students. The *PSR* also provided similar constructs that addressed reading fluency and comprehension skills and was designed for interval seasonal administrations. Roberts, Good, & Corcoran (2005) found a high correlation coefficient of .76 for third grade students who were tested using oral reading fluency measures to possess grade level or above reading and comprehension skills.

The *PSR* was a proprietary test that gave an overall score that was influenced by similar constructs as the *DORF*. Messick (1989) and Burns (2007) concluded that developing methods that assessed how decisions that were generated using multiple measures were essential to an accurate evaluation practice. Further consideration for construct validity must be supported through testing the accuracy of the test interpretation of a given model (Kane, 2001; Salkind, 2012). Both the *DORF* and *PSR* have ample reliability and validity studies and research that supported the identification of student reading performance. Understanding how the PM measures work together to assess the underlying constructs of a LD profile was imperative. The combined use of the measures was used to identify a LD profile when the data derived from the PM measures were utilized as a tracking tool rather than a diagnostic instrument. Burns asserted that further research was indicated to support the interpretation of data profiles in order to consistently interpret RTI data for the use of LD diagnosis.

Learning Disability Assessment of Diverse Lower Socioeconomic Populations

In order to identify a score profile for students with LD, a multi-tiered screening process was recommended (United States Department of Education, 2013). No specific language in the legislation was provided to address diversity of culture, race, or lower socioeconomic populations of students. The study population for the current research project, presented with numerous concerns about the lower scores that were evidenced on the universal screening and PM measures when compared to the national norms used by the instruments. Decisions about intervention instruction and potential evaluation considerations were acknowledged through local norming.

Lower socioeconomic status was defined as communities that presented with individuals and families that were dependent on government subsidies and programs. Based on the poverty threshold income data supported by the United States Census Bureau (2012), the annual wages for 2009 was $11, 720 for an individual and $17, 603 based on a family of four. The average yearly income of US residents, according to the US Census Bureau reports from July 1, 2000 records, was $53, 048. The 2012 United States Census Bureau reported that 30.9% single parent families reported living below poverty levels in Illinois.

Of the 12.4 million Illinois residents, 3.8 million family members were impacted by poverty level incomes. Included in this demographic data, a quarter of African-Americans and Hispanics in the United States were identified by the United States Census Bureau (2012) as at or below the poverty level. In the community where the current study was conducted, 86% of the children came from poverty level conditions. These students were impacted by lower socioeconomic environmental conditions that

attended the district schools. These environmental conditions all presented with circumstantial problems that interfered with disproportionality concerns related to over identification of special education disabilities (Illinois State Board of Education, 2012b).

The MTSS decision-making process required a consistent set of contingencies based on local academic performance instead of comparing the students' scores to national norms. Local norming of universal screening measures provided the data that guided curriculum-based interventions and decisions in these circumstances of diverse socioeconomic communities (Stewart & Silberglitt, 2008; Villegas & Irvine, 2010). Students who were not responding to interventions, despite numerous attempts to address the needs of the child in the general education intervention process, required more intensive interventions through special education services (Sibley et al., 2001). The data profile would determine whether further support was necessary for the children who were struggling academically or who may have LD.

Locally-normed PM measures that addressed the academic growth utilized to monitor the students from diverse communities could provide the essential eligibility data that supported decisions for LD in accordance with Individuals with Disabilities Improvement Act legislation (United States Department of Education, 2013). Comparing the students to their peers through the use of local norms was the logical practice to consider. This process may also prove to be a reliable manner to avoid the criticism to the MTSS of what was considered as *watching* the student fail until the students met the criteria for further intervention or ultimately for a referral for special education evaluation (Mellard et al., 2010; Reynolds, 2008).

Burns (2007) addressed how the assessment paradigm shifted from the assessment of student learning patterns to an evaluation of how the child learns. Research conducted with minority children reflected the impact on learning based on cultural, socioeconomic, and linguistic factors. Often, these minority children were over-identified if not misidentified with LD when other factors were influencing their learning. This misidentification reflected a disproportionality of LD diagnoses when compared to their Caucasian peers (Compton et al., 2010; Pearce & Gayle, 2009; Sullivan et al., 2009; Tyler, Yzquierdo, Lopez-Reyna, & Flippin, 2004).

Conclusion

The implications of cultural diversity in the public schools required programming and assessment modifications that identified and supported the academic needs of a diverse population of students (Villegas & Irvine, 2010). Utilization of local norms to support the minimization of the misidentification of LD with students who were exposed to environmental, linguistic, and social conditions appears as a logical intervention was noted as important (O'Connor & Klingner, 2010; Skiba et al., 2005; Utley et al., 2011). Students with diverse backgrounds may hold a higher risk of disproportionality of special education services and may not benefit from the direct comparison to the national population norming practices. Therefore, the utilization of local norms, the necessary corrective instructional interventions, and supports to children with cultural diversity could reduce the number of children misidentified with LD.

Summary

Learning disabilities present with numerous factors that make evaluation and accurate diagnosis a challenging process. The use of MTSS practices for the assessment

of children with reading problems that indicated a link to learning disabilities presents with several factors that should be considered when using universal screening benchmark PM scores as the assessment data. Using universal screening methods that included proprietary tests and curriculum-based measures are required to show a student's lack of response to MTSS. Proprietary PM evaluations do not offer information about how eligibility decisions were made other than to produce the nationally-normed cut-off scores that provided suggested recommendations for further interventions. The misidentification of LD of students in a culturally, socioeconomic, and linguistically diverse school population using national norms often resulted in a Type I Error where the child was incorrectly diagnosed with a learning disability.

The use of local norms to assess diverse lower socioeconomically impacted student populations was suggested as a means to quantify the differences of children with LD when compared to grade-related peers within the curriculum used by the school. The evaluation process required the use of two reading-based measures that reported scores with different statistical properties to empirically verify a child's lack of academic growth based on the intervention. The use of *Z*-score standardization facilitated the comparison of these measures. The utilization of slope methods and discrepant scores supported the academic progress documentation. The utilization of local norms for students provided the comparison to address the identification of the children with LD. Comparing this data through the use of standardized scores for the assessment of LD may provide a more consistent model for identifying children through the use of local norms across measures.

CHAPTER III

METHODOLOGY

Introduction

The current study focused on concerns of the misidentification of SLD in reading with children living in predominantly lower socioeconomic communities. Several factors were addressed in this research regarding the use of national norms that assessed the children who have been impacted by environmental, socio-economic, and cultural factors. The use of Multi-Tiered Systems of Support (MTSS), also referred to as RTI, and the means to evaluate the children through universal screening relied on national norms in order to assess student academic progress within a given curriculum.

The universal screening scores that were reflective of the children in lower socioeconomic communities were often lower than the national norms. In these circumstances, the children were frequently misdiagnosed with learning disabilities (Sullivan & Artilles, 2011; Utley et al., 2011). In order to limit the misdiagnosis of SLD, local norms were considered that reflected the student distribution of the school district and learning needs. These concerns were based on how the community impacted environmental, economic, cultural, and familial conditions that influenced student learning patterns (Skiba et al., 2005). The research design and analysis of the current study examined the demographics of the population, universal screening methods used by the school district, the analytic methods used to evaluate the research questions, and

limitations of the study. Moreover, the statistical methods used to assess both the actual and inferential findings were addressed.

Research Design

The universal screeners that were utilized for the current study focused on the reading skills development of third grade students. Reading performance data was generated from PM scores based on two forms of universal screening tests that were proprietary. The *Dynamic Indicators of Basic Early Literacy (DIBELS) Oral Reading Fluency* (*DORF*) and the *Performance Series Reading* (*PSR*) scale scores were the two forms of reading evaluation instruments utilized. These evaluation measures were used as universal screeners to assess students' response to the reading curriculum within the school district studied.

The two reading measures that were utilized, the *DORF* and the *PSR,* offered reliable and valid scores for monitoring student academic performance in reading on a continual interval basis (Briggs, 2011; Compton, Fuchs, Fuchs, & Bryant 2006; Cummings, Kennedy, Otterstedt, Baker, & Kame'enui, 2011; Kim et al., 2010; Shaw & Shaw, 2002). Both of these measures were used extensively by schools throughout the United States. The extensive use of the PM measures reflected the general consensus of assessment teams that the tests provided a reliable means for obtaining scores for the routine monitoring of student progress in reading in the elementary through the high school settings (Jenkins et al., 2013; Scott & Weishaar, 2003; Stecker et al., 2008).

The PM measures used in the current study provided technical and/or psychometric properties that adequately reflected reliability and validity (Deno, Mirkin, & Chaing, 1982; Utley, 2011). The achieved PM scores reflected the reading

performance over time. The student data provided an effective means to communicate evaluative judgments regarding student progress for intervention efficacy and/or LD assessment (Hagermoser-Sanetti et al., 2011; Stecker et al., 2008). The data that was generated from the seasonal universal screening scores allowed for sensitivity to the changes in the performance of the child (Ardoin & Christ, 2009; Faggella-Luby & Wardwell, 2011). The RTI data was essential for assessment information for screening and eligibility purposes (Jenkins et al., 2013; Tran, et al., 2010).

The seasonal universal screening that occurred three times throughout the school year provided the data for comparison of the third grade students' performance. The data were collected during the school year 2012-13. The evaluations administered were first conducted approximately 15 days into beginning of the year. The next round of universal screening for reading occurred at the 90-day mark (i.e., 75 days after the initial assessment), which was referred to as the middle of the year evaluation. The last interval evaluation occurred 75 days after the midterm and was referred to as end of the year data. MOY data was not used in the current study. The focus of the study was to examine the students' growth over the entire school year.

The student scores were arranged based on student ID and the screening instruments' interval test periods. Data from each test date were analyzed separately, and then compared to the actual number of students identified with learning disabilities in reading based on local and national norms. The mean and standard deviations were calculated for the participants in order to establish local norms. With this information, the standard scores were then computed and utilized for comparison.

There were several types of standardized scores considered that provided an approach to compare and analyze data. For the current study, *Z*-scores were utilized. The scale scores of the *PSR* and the raw scores of the *DORF* measures were converted to *Z*-scores. This conversion to standard scores was essential to the comparison and classification of the scores. The *Z*-scores calculation required the datum of one student minus the mean of the sample divided by the distribution standard deviation (Salkind, 2012). The formula to complete the conversion of the individual's scores was as follows:

$$z = \frac{(\times - \bar{x})}{s}$$

Where:

z was the *Z*-score computed

$\times$ was the participant's individual score

$\bar{x}$ was the mean of the sample

s was the standard deviation of the sample

The mean of *Z*-scores is zero and the standard deviation is one. A positive number is indicative of a score above the mean while a negative number reflected a score below the mean.

The individual samples of each instrument's scores were standardized. This standardization allowed for the comparison and statistical manipulation of the scores, and the numerical classifications of the data were easily identified. The classifications of a student's scores were compared based on the cut-off criteria of the *Z*-score of -1.27 for both of the national and local norms. The *Z*-score of -1.27 was consistent with the 10^{th} percentile, which was identified in the research literature as an appropriate cut-off score to further assess for learning disabilities (Shinn, 2007). The evaluation of national and

local norms of the children who were classified as SLD in reading was conducted by comparing the nominal data through the use of Chi-Square, using those *Z*-scores as the method to classify at-risk students. The nominal scores were reflected as *YES* for the students' scores that met the cut off criteria of a -1.27 *Z*-score and lower, and *NO* for the children who scored above the cut-off score. These statistical processes were the foundation methodology used to answer the following research questions:

1. What differences exist in the identification of specific learning disabilities when using local norms classification compared to national norms classification?
2. What resulting cut-off score is determined when utilizing slope based on locally-normed data?
3. How was the accuracy of the identification of specific learning disabilities affected when both performance scores and slope differential were used?

Population

The population consisted of 428 third-grade children in a Northeastern Illinois school district. The learning disabilities that were addressed in the current study focused on reading skills performance data generated from PM scores of third-grade children. These third graders attended a total of five separate elementary schools within the school district. The student population during school year 2011-12 consisted of 51% African-American, 24% Hispanic, 20% Caucasian, and 5% other. At least 84% of the student population came from low-income families and received federally supported funding for breakfast and lunch within the school district. Of the 428 students, 47 of the third-grade students were diagnosed with SLD prior to the current study. The 47 students, who

comprised approximately 11% of the sample, were identified by a diagnosis of SLD and had an Individualized Education Plan (IEP) with reading goals.

Data Collection

Answering these research questions required a system that was utilized within the school district that routinely assessed children at marked intervals throughout the year. The *DORF* and the *PSR* were administered three times during the year. Based on the current study, the *beginning of the year* and *end of the year* scores were utilized with an elapsed 170 school days between administrations. The MOY scores were not used, in order to reflect a full school years' worth of growth. These tests were conducted within three weeks or 15 days into the *beginning of the year* and within three weeks or 15 days before the *end of the year*. With this system established, the third-grade student scores generated for the entire school year were reviewed and analyzed. Specific to the instruments used, the means and standard deviations were calculated for the local data. Similarly, the mean and standard deviations for the nationally-normed data were used to examine the score differences.

The DIBELS Oral Reading Fluency Measure

The *DORF* was used to measure reading fluency and comprehension. The measure was administered to each of the students individually by a trained teacher who was not the child's teacher. The test was scored by counting the number of words read by the student within the passage that occurred during a one-minute timeframe. The Cronbach's Alpha Coefficient that provided the internal consistency of the *DORF* was .91. Correlation coefficients that were rated between the three seasonal administrations of third graders ranged between .89 and .93 (Shaw & Shaw, 2002). The standard deviation

(*SD*) based on the national norms for the beginning of the year was 35.91 with a mean of 79.89, middle of the year *SD* was 37.09 with a mean of 96.30, and the end of the year screening possessed a *SD* of 38.26 with a mean of 107.20. Other technical characteristics of the *DORF* reflected test-retest coefficients of .92 to .97. Based on criterion-related studies that were conducted, the *DORF* coefficients ranged between .52 to .91. The *DORF* was a reliable tool for screening children for their reading skills development.

The nationally-normed data was broken down into classifications that identified the level of risk for reading difficulties. There were score ranges for three categories of risk. The categories consisted of at risk, some risk, and low risk. Based on the time of year that these students were assessed, specific performance ranges were provided by the authors. For example, students who fell in the *at risk* category in the fall of the school year had a performance raw score on the *DORF* of less than 53 words per minute (wpm), which required intensive interventions. Children with *some risk* fell within the performance category of greater than or equal to 53 to less than 77 wpm and were supported through strategic interventions. Within the same testing period, students who fell in the *low risk* classification performed at 77 wpm are better and fell within the reading development expected range (Good et al., 2004).

The Performance Series Reading Scale

The second test administered periodically throughout the third-grade school year was the *PSR*. The Cronbach's Alpha Coefficient that addressed the internal consistency of the *PSR* national norms was .91 and reflected a mean Standard Error of Measurement of .29. A 12-week, or at least 60-school day, window of time between test administrations

was recommended based on the nationally-normed recommendations for testing. The administration time was approximately 30 minutes for the reading test.

The *PSR* universal screener was a computer-administered test. The student was presented with a reading passage that was generated online. The passage level we based on the student's grade level and time of year. Once the student completed reading the passage, they were prompted to answer multiple-choice questions. The questions were based on the child's responses. If the child did not respond correctly to the first answer, the content was decreased in complexity based on the child's responses. If the student attempted to start the questions within a given timeframe that was not reasonable for completing the passage, the program would spoil the student's test.

The several different scale scores that were available on the *PSR* were based on the Rasch Model and ranged from 1300 to 3700. The Rasch Model was based on probability measurement models (Bond & Fox, 2012). This model was used mainly for testing that indicated a development or lack of development of skills or knowledge over time. The scale scores were based on the continuum and were used as a means for comparison between peers based on grade level reading norms.

The reading scale scores were available on a separate website after the child completed the assessment. These scores were accessed by the student's teacher and other approved school personnel online. Instructional suggestions and intervention goals supported the responses for each of the children tested. The scale scores included reading rate, vocabulary, Lexile (proprietary reading level score) level, grade-level estimate, fiction and nonfiction scores, and an overall reading scale score. The reading scale score

was utilized for the current study. The school district average score and sample number were provided as a comparison for each time the student completed the periodic screener.

The *PSR* nationally-normed sample of third grade students was based on a 50th percentile score of the three testing periods separately. The *beginning of the year* scale score was 2298, the MOY scale score was 2396, and the *end of the year* scale score was 2467. All of the mentioned scores were based on the national norms with a standard error of measurement of ± 34 and a *SD* of 159.05. The mean gain of a third grader that occurred between the first and last interval test dates within the school year was 212 points. For children who fell between the national percentile rankings of 41 and 50, the *PSR* scale scores that ranged from 2204 to 2298. This score range equated to a targeted reading growth of 287 points during a school year. The growth target fluctuates based on the national percentile ranking. The *PSR* scale scores focused on the change in a student's scale score progress over time. A score below or above the target growth rate of 287 points between the fall and spring testing was the manner, which the scores were compared. The categories were based on the child's growth performance as either *far below, below, above, or far above*. The complexity of the scale score system was another reason for converting these scores to *Z*-scores and developing local norms in order to identify children who struggled with reading skills.

The variables were identified as the collected universal screening scores of the third-graders for the three seasonal administrations using both the *DORF* and the *PSR*. The at-risk for learning disabilities scores of the students based on local classifications were the scores of the children who were below the 10th percentile, or *Z*-score of -1.27 (Christ et al., 2012; Good et al., 2001; Shinn, 2007). This score was based on the number

of children within the population who were identified with learning disabilities that were identified by the state of Illinois based on a 2012 audit of students with special education eligibilities. The national average for students with learning disabilities was identified between 4% and 9% (American Psychiatric Association, 2013; Centers for Disease Control, 2011; Illinois State Board of Education, 2009). The school district had identified 6.7% of the student population who were diagnosed with specific learning disabilities, mainly in reading, using the ability-achievement evaluation model in 2012.

There were two levels of variables that were compared between the national classifications of students with potential learning disabilities and the local norms that were being used to identify children within the school district. The variables were based on the nationally and locally-normed *beginning of the year* and *end of the year* reading scores for *PSR* and *DORF*. Two levels of data for each of the tests, based on the scores from the September and May universal reading screening administrations, were used to identify student yearly progress.

The next task concentrated on the cut-off points based on the data for the students who were reading below expectations. In order to convert the interval data to nominal data, *Z*-scores were computed to identify the cut-off score for both the *PSR* scale scores and the *DORF* raw scores. The nationally-normed instrument's cut-off scores were based on the -1.27 *Z*-score, which was equivalent to the 10^{th} percentile.

Based on the nationally-normed data assessment protocol, children who fell below one standard deviation below the mean, or 15^{th} percentile, were considered for further assessment needs. This level of cut-off score was identifying an excess number of students who potentially had SLD in the school district that was studied. The research

that supported further assessment of children who were identified to be *at risk* for reading disabilities fell in the range below the 10th percentile, which equates to the *Z*-score of -1.27 on local norms (Christ et al., 2012; Good et al., 2001; Shinn, 2007). The utilization of local norms was considered as a potential support to reduce the number of children who were identified using national norms. Local norms were developed based on the scores of the school district's third-grade children for each of the benchmark, or interval screening, administrations.

Analytical Methods

Research Question Number One Methodology. What differences exist in the identification of specific learning disabilities when using local norms classification compared to national norms classification?

The accurate identification and classification of the interval testing throughout the school year for each student was based on research by Shinn (2007). This research addressed cut-off scores based on percentile. The recommended cut-off scores for SLD were between the 5th and 10th percentile, according to Shinn. Because percentiles were not designed as a comparison among other scores, but rather a comparison among members of a sample group, percentiles were not used for this purpose. Instead, the use of standardized *Z*-scores based on the equivalent score of the 10th percentile, or a *Z*-score of -1.27, was utilized in the current study to support the classification process for the comparison of the scores based on local and national norms. Any scores at or below a -1.27 *Z*-score were indicated as a potential classification for SLD. With this fundamental information, the comparison of the students with scores below the 10th percentile, or a *Z*-score equal to or below -1.27 based on local norms, were utilized for the classification

process. The Chi-Square statistic was used to identify the differences of the children who fell below the 10th percentile for the locally and nationally-normed data for both the PRS and *DORF*. Descriptive statistics for the local, national, and diagnostic information were included in order to address the frequencies of the data.

The variables were the classification of the identified children with scores that fell below the *Z*-score of -1.27 based on the mean of the test scores on the local and national norms. The classifications based on the national norms of children with reading deficiencies were also converted to *Z*-scores. Classifications for the nationally-normed data were then based on a *Z*-score of -1.27, or 10^{th} percentile, and lower for the given universal screening measures used in the school district.

The 47 students with diagnosed SLD were identified within the data groups for both the *PSR* and *DORF*. These 47 students were identified with SLD using the ability-achievement methodology. The diagnoses of SLD occurred prior to 2012 when the current study was initiated. However, due to the high level of transiency some of these students did not take the *end of the year* benchmark testing. Therefore, the data does not reflect the original number of students identified with SLD. The children were found through a record search of who had an individual education plan (IEP) for learning disability eligibilities in reading. The classification process for both the national norms and local norms was essential to help with the comparison accuracy of the categorical data with the diagnostic eligibility classification of SLD.

Three classifications were used to represent the differences between the cut-off score and the actual identification of the students with SLD in reading. These classifications included: Diagnosed SLD (those formally diagnosed with a learning

disability in reading), national SLD (those who fell below the national norm 10th percentile on the *PSR* and/or *DORF*, indicating that the student may have a SLD), and local SLD (those who fell below a *Z*-score of -1.27 on the locally-normed data, which corresponds with the 10th percentile).

Chi Square was used to determine whether the classifications were different than the identified children with a prior diagnosis of SLD. Chi Square was used for the *beginning of the year* and *end of the year* data for the *PSR* and *DORF* scores for the local and national norms.

Research Question Number Two Methodology. What resulting cut-off score is determined when utilizing slope based on locally-normed data?

The Tukey slope model was utilized to assess change over time. This form of slope was determined by subtracting the first assessment from the second assessment and dividing by the number of elapsed days based on the most recent assessment minus the number of days that were elapsed from the first administration as follows:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

Where:

m is the individual slope coefficient

y_1 first score of the universal reading series

y_2 last score of the universal reading series

x_1 first assessment administration day count 5

x_2 last assessment administration day count 175

The individual slope scores were based on the first and last administration of the *DORF* and *PSR*, with the *beginning of the year* score subtracted from the *end of the year*

PM score. The first administration fell on an average of the 15th day into the new school year and the last evaluations were conducted on approximately the 165th day of the 180 total days of the school year. For the current study a 170-day lapse of schools days was used to divide into the numerator, which was based on the subtraction of the *end of the year* score from the *beginning of the year* score of each student. Once the individual slopes were computed, the mean of the slope coefficients were calculated. The group slope coefficients were summated and divided by the number of student participants in order to calculate the mean. This process was completed based on the local and national norms. At this point, the calculations of the mean of the slope scores, standard deviation, and *Z*-scores were computed for each of the participants' slope scores. As described with the discrepancy scores, the cut-score for the slope was a *Z*-score -1.27 or lower. The slope score was used to identify the participants who were not progressing in reading in a manner consistent with their peers.

Research Question Number Three Methodology. How was the accuracy of the identification of specific learning disabilities affected when both performance scores and slope differential were used?

The third question addressed the accuracy of identification of SLD when both performance and slope differentials were utilized as diagnostic information. The variables examined were the classification of a student with poor performance in reading based on local norms only versus the classification of the same student based on local performance norms and local slope differential norms that provided conditions to correlate.

The final research question was addressed through the correlation of the discrepancy cut-off scores and the slope scores. This allowed for the examination of the

relationship of these two variables with the diagnosis of reading-based learning disabilities. The next step focused on the determination of whether the local scores and slopes improved the identification of a student's specific learning disability in reading. The results of this data were compared to the children who actually had IEPs for reading disabilities through nonparametric correlation statistics.

The classifications based on the *PSR* and the *DORF* scores were further explored. Each of the universal screening measure scores were based on the locally-normed data. The students' reading performance scores and slope coefficients were identified as either *YES* he or she did meet the criteria that the scores that were equal to or below a *Z*-score of -1.27 or *NO* he or she did not. The discrepant students based on the *PSR* and *DORF* scores were compared individually to the students who were diagnosed with SLD. Then the combination of the scores and the slope coefficients were correlated using the nominal data. The nominal data considered of *YES* the score met the cut-off criteria and *NO* the score was above the cut-off criteria.

Both variables were based on nominal data. The use of the Contingency Coefficient was utilized in order to assess the correlation of the nominal scores and slope coefficients of the locally-normed data.

Limitations

Several limitations were evident based on the evaluation criteria for specific learning disabilities. Although the assessment of SLD was outlined in Individuals with Disabilities Educational Improvement Act, the law did not provide specific evaluation recommendations. The legislative guidelines were based on how well a child responded to interventions over time as well as other normed-referenced evaluations such as

achievement and intelligence testing. Concerning the degree that the students responded to interventions that was indicated by their performance on PM measures, no definitive statements or suggestions were provided by the law makers. The local norms developed were based on the research sample to quantify and address the lack of guidance proposed by the legislation.

The students who were actually diagnosed with SLD were evaluated and found eligible for special education services through the method known as the ability-achievement method. The results of the standardized tests presented with a statistically significant discrepancy between the IQ and the norm-referenced reading achievement score/s. The IQ of the students that were diagnosed with SLD fell in the average range of intellectual functioning. The discrepancy was based on national norms and was at least one standard deviation or greater below their age- and/or grade-related peers on the reading scores. This form of evaluation method used normed reference measures and utilized a traditional psycho-educational battery. No benchmark or progress monitoring data was utilized for the classification or diagnosis of SLD for this comparison group in the current study. The current study provided an investigation of the possible outcomes of the use of local norms with classification of SLD utilizing local norms for benchmark testing.

A further limitation of the current study was associated with the inclusion of only the third graders in the school district and not the full spectrum of the other grades. The results of the current study were used as inferential information only. Further evaluation was recommended to address other grade profiles related to local norms as well as the measures used to screen for reading problems.

An additional concern was that after the third grade, the *DORF* was not used. The only evaluation measure that provided benchmark-testing results beyond the third grade was the *PSR*. Other forms of fluency-related screening tools are available commercially. Other forms of universal screening measures were recommended to support the seasonal reading assessment of the students based on the guidelines of Individuals with Disabilities Educational Improvement Act. Therefore, the use for the *DORF* measure after the third grade was not utilized. By not using the *DORF* after the third grade, the implications from the current research to successive grades presented with a limitation and cautionary note to the evaluation team. The data derived from the *DORF* was instead used as a historical marker that reflected past performance that could be reviewed when the student was reevaluated at a later time.

Another limitation was that the data from year to year changed for each grade set of students. Performance monitoring scores fluctuate from class to class and year to year. These changes were based on intended and unintended changes to the curriculum and instruction of the schools. When local norms are updated, the score fluctuation will impact the other criteria that included discrepancy cut-off scores and slope, which will impact potential diagnostic data for specific learning disabilities.

The community where the study was conducted was highly mobile. Students would move to and from the school district on a regular basis. This movement was indicative of families being evicted from their homes, homelessness, parents being incarcerated, foster child placement, and other extenuating and mitigating circumstances that were relevant to lower socioeconomic communities. As a result, scores were not

always available that corresponded with either the *beginning of the year* or *end of the year*. Transiency of families was the main factor for this limitation.

One of the overlooked limitations was the complexity of the *Z*-score system that was not understood by the other school districts that received the children with learning disabilities. The receiving schools were not aware of the specific criteria process for the diagnosis of SLD that the study school district used. In order to minimize this problem, careful consideration of the evaluation documentation was necessary. Specifically, the documentation addressed the use of local norms and use of percentiles for the diagnosis process helped to eliminate these diagnostic questions.

The accuracy of the diagnosis of a SLD in reading of the 47 students qualified as a limitation as well. The accuracy of the diagnosis of this group was based on the judgment of the assessor and the data interpretation based on the ability-achievement model. Based on the locally-normed data, the number of students with potential or actual SLD who were identified was conservative, based on the 10^{th} percentile cut-off score. The minimization of false negatives was potentially compromised due to missing student data.

Summary

The purpose of this chapter was to provide a description of the methodology to address the utility of locally-normed universal screening measure scores. The population consisted of 428 third grade students who lived in a lower socioeconomic community. A minimum of 84%, or approximately 360 of these third graders, came from families that lived at or below poverty levels established by the United States Census Bureau (2012). The universal data was derived from the *PSR* and *DORF* measures that were

administered at the *beginning of the year* and *end of the year*. The scores were converted to standardized *Z*-scores in order to differentiate them for classification. A *Z*-score of -1.27, or 10th percentile, was used as the cut-off score for the classification of SLD for both national and local norms. Chi-square was used to find the differences of the local and national-normed results to identify the discrepant scores. Slope scores were computed to assess student progress and correlated with the discrepancy scores for further utility in the classification process of learning disabilities. The limitation of the study were discussed, including legislative guidelines, accuracy of the diagnosis of SLD, yearly data considerations, and the use of a conservative cut-off score for the classification of the data. The findings, conclusions, implications, and recommendations of the current study will be further explored in the next chapter.

CHAPTER IV

FINDINGS AND CONCLUSIONS

Introduction

The premise of the current study was based on the requirements of Individuals with Disabilities Educational Improvement Act for the intervention and evaluation process of diagnosing reading learning disabilities. The changes in the legislation required the evaluator to use multiple methods to evaluate SLD. Historically, the administration of universal screening in reading was used to assess students' performance responses based on results of the interventions. This universal screening method was known as CBM. With the onset of the new IDEA regulations, proprietary screening measures were developed. The various measures reported student progress in an array of scores that did not readily compare to one another to support the evaluation process of SLD. The scores of the reading measures were normed utilizing national norms. This practice often led to the misidentification of students with learning disabilities who came from socially, economically, and culturally diverse communities (Utley et al., 2011).

The current study explored whether scores based on local versus national norms minimized the misidentification of learning disabilities in reading for minority student populations. Based upon previously-conducted research, nationally-normed reading measures tended to reflect lower scores for individuals from diverse populations (Tran, et al., 2010; Utley, 2011; Yeo et al., 2011). These scores could lead the assessor to

misidentify reading learning disabilities due to the nationally-normed cut-off score recommendations of the instruments. Slope data were used in the current study in conjunction with the two reading measures scores to determine any correlations in an effort to further support a lack of academic growth. The standardized scores were compared across measures. Overall, the methodology of the current study provided a method of limiting Type I Errors or over-identification of SLD in reading with children from culturally and environmentally diverse communities.

The research questions answered in the current study addressed the need for the consideration of local norms when specific learning disabilities were over-, under-, and misidentified in communities with diverse cultural and lower socioeconomic environmental conditions (Sullivan & Artiles, 2011). The utilization of two forms of reading PM data in conjunction with slope coefficients offer several forms of evaluation data for a conventional method of diagnosing learning disabilities (United States Department of Education, 2013). The research questions addressed were:

1. What differences exist in the identification of specific learning disabilities when using local norms classification compared to national norms classification?
2. What resulting cut-off score is determined when utilizing slope based on locally-normed data?
3. How was the accuracy of the identification of specific learning disabilities affected when both performance scores and slope differential were used?

Findings

The students who were diagnosed with SLD consistently performed at a lower rate than their peers. There was a 219.49 point difference between the *beginning of the year PSR* scores of general education students and the children with SLD. *PSR* scores at the *end of the year* reflected a 330.05 difference between the SLD and general education students. In both of the score profiles from the *PSR* based on *beginning of the year* and *end of the year*, the students without disabilities scored higher (Table 1).

Table 1

Performance Series Local Scores Descriptive Statistics

PSR Classification	BOY Mean (*SD*)	EOY Mean (*SD*)
Diagnosis of SLD (*n* = 47)	1922.97 (244.39)	2016.84 (304.80)
No Diagnosis of SLD (*n* = 352)	2142.46 (242.56)	2346.89 (252.89)

A similar pattern of scores were found from the *DORF* raw scores. Children with the diagnosis of SLD scored lower than the general education population. The difference between the students diagnosed with SLD and the students without SLD *beginning of the year* scores was 36.77. This pattern of discrepancy was found at the end of the year, but to a greater degree, with a 48.61 word per minute difference in the scores between the two groups. Therefore, fluency rates for children with SLD in the third grade appeared to be well below that of their peers (Table 2).

Table 2

DIBELS Oral Reading Fluency Local Scores Descriptive Statistics

DORF Classifications	BOY Mean (*SD*)	EOY Mean (*SD*)
Diagnosis of SLD (n = 42)	36.73 (27.79)	49.88 (37.86)
No Diagnosis of SLD (n = 334)	73.50 (31.01)	98.49 (31.01)

Research Question Number One. The first research question probed the universal screening measures' score differences of national and local norms classification of SLD. The *PSR* and *DORF* descriptive statistics were derived from the *beginning of the year* and *end of the year* scores. As previously noted, students received a classification within each of the following categories: (a) Diagnosed SLD, (b) National SLD, and (c) Local SLD.

In order to determine if the local norms outperformed the national norms in the correct identification of students diagnosed with a learning disability in reading, Chi-Square analyses were conducted for both the *PSR* and *DORF* across the two assessment periods, *beginning of the year* and *end of the year*.

When the 10^{th} percentile is utilized (i.e., *Z*-score -1.27), the *beginning of the year PSR* national norms indicated that 49.2% (or 178) of all students met the classification criteria of SLD (Table 3). This assessment correctly classified 81.1 % (or 30) of the students who were actually diagnosed with an SLD. This difference was statistically significant (X^2 (1, N = 362) = 16.79, p < .001). Of particular note, the national norms erroneously classified 45.5% (or 148) of the students as having an SLD when they did not.

Table 3

Performance Series Reading BOY SLD Categorization Utilizing National Norms

Classification	Local False	Local SLD
Diagnosed False	177 (54.5%)	148 (45.5%)
Diagnosed SLD	7 (18.9%)	178 (49.2%)
Total	184 (50.8%)	178 (49.2%)

A statistically significant difference was also found when *beginning of the year PSR* local norms were compared with the diagnosis of a reading-based learning disability (X^2 (1, N = 362) = 24.32, $p < .001$). The percentage of diagnosed SLD children correctly classified with the local norms was only 35.1% (or 13 students) (Table 4). These data indicated that 64.9% (or 24) of students with an actual diagnosis of SLD were not properly classified, which reflected an under-identification of the classification.

Table 4

Performance Series Reading EOY SLD Categorization Utilizing Local Norms

Classification	Local False	Local SLD
Diagnosed False	298 (91.7%)	27 (8.3%)
Diagnosed SLD	24 (64.9%)	13 (35.1%)
Total	322 (89.0%)	40 (11.1%)

The *PSR end of the year* national norms classified 42.2% (or 157) of the students as likely to have SLD based on scores below the cut-off *Z*-scores -1.27 (Table 5). Of the 332 students who were not diagnosed with a SLD, 38.0% (or 126) of these children, were classified as SLD when the national norms were utilized. The difference in classification

was statistically significant (X^2 (1, N = 372) = 22.89, p < .001). The national norms classified 22.5% (or 9 students) as not having SLD when they did have the diagnosis.

Table 5

Performance Series Reading EOY SLD Categorization Utilizing National Norms

Classification	National False	National SLD
Diagnosed False	206 (62.0%)	126 (38.0%)
Diagnosed SLD	9 (22.5%)	31 (77.5%)
Total	215 (57.8%)	157 (42.2%)

A statistically significant difference was also found when *end of the year PSR* local norms were compared with the diagnosis of a reading based learning disability (X^2 (1, N = 372) = 34.93, p < .001), much fewer children were misclassified as learning disabled 9.3% (or 31 students) (Table 6). With use of local norms, 57.5% (or 23) of students with an actual diagnosis of SLD, were not properly classified, supporting under-identification.

Table 6

Performance Series Reading EOY SLD Categorization Utilizing Local Norms

Classification	Local False	Local SLD
Diagnosed False	301 (90.7%)	31 (9.3%)
Diagnosed SLD	23 (57.5%)	17 (42.5%)
Total	324 (87.1%)	48 (12.9%)

A statistically significant difference was also found when *beginning of the year DORF* national norms were compared with the diagnosis of a reading-based learning disability (X^2 (1, N = 380) = 79.00, p < .001). Of those diagnosed with a SLD, 60.0% (or 24 students) were also correctly identified through use of the classification based on the national norms (Table 7). However, with the use of national norms, 40.0% (or 16) of students with an actual diagnosis of SLD were not properly classified, suggesting a possible under-identification.

Table 7

DIBELS Oral Reading Fluency BOY SLD Categorization Utilizing National Norms

Classification	Local False	Local SLD
Diagnosed False	311 (91.5%)	29 (8.5%)
Diagnosed SLD	16 (40.0%)	24 (60.0%)
Total	327 (86.1%)	53 (13.9%)

A statistically significant difference was also found when *DORF end of the year* local norms were compared with the diagnosis of a reading-based learning disability (X^2 (1, N = 380) = 82.33, p < .001). The percentage of children who fell in the Local SLD category that were not diagnosed with a SLD was approximately 6.5% (22 students). However, with the use of local norms, 45.0% (or 18) of students with an actual diagnosis of SLD were not properly classified, suggesting a possible under-identification (Table 8).

Table 8

DIBELS Oral Reading Fluency EOY SLD Categorization Utilizing Local Norms

Classification	Local False	Local SLD
Diagnosed False	318 (93.5%)	22 (6.5%)
Diagnosed SLD	18 (45.0%)	24 (55.0%)
Total	366 (88.4%)	44 (11.6%)

The *DORF end of the year* national norms classified 18.0% (or 66) of the students as likely to have SLD based on scores below the cut-off *Z*-scores -1.27 (Table 9). This classification of SLD within the general education sample was in contrast to the 69.0% (or 29) of the students who were actually diagnosed with SLD that were classified with national norms. The difference was found to be statistically significant (X^2 (1, N = 366) = 83.54, $p < .001$). The national norms classified 31.0% (or 13 students) as not having SLD when they actually had the diagnosis.

Table 9

DIBELS Oral Reading Fluency EOY SLD Categorization Utilizing National Norms

Classification	National False	National SLD
Diagnosed False	287 (88.6%)	37 (11.4%)
Diagnosed SLD	13 (31.0%)	29 (69.0%)
Total	300 (82.2%)	66 (18.0%)

A statistically significant difference was also found with the *DORF end of the year* scores when local norms were compared with the diagnosis of a reading-based learning disability (X^2 (1, N = 366) = 90.50, $p < .001$). The percentage of children who fell in the Local SLD category was commensurate with total of the diagnosed sample, which was 11.2% (or 41 students) (Table 10). However, with the use of local norms, 42.5% (or 19) of the students with an actual diagnosis of SLD were not properly classified, that suggested a possible under-identification.

Table 10

DIBELS Oral Reading Fluency EOY SLD Categorization Utilizing Local Norms

Classification	Local False	Local SLD
Diagnosed False	306 (94.4%)	18 (5.6%)
Diagnosed SLD	19 (45.2%)	23 (54.8%)
Total	325 (88.8%)	41 (11.2%)

Research Question Number Two. The second research question was based on the identification of a cut-off slope score coefficient derived from utilizing a -1.27 *Z*-score, or 10^{th} percentile. Annual growth slope was calculated from the *beginning of the year* and *end of the year* benchmark testing using the Tukey method to determine slope scores. Individual student slope scores were calculated by subtracting the end of the year from the beginning of the year scores and dividing by the 170 days that had lapsed between the *beginning of the year* and *end of the year* administrations of the reading test. This

calculation provided the coefficient scores that yielded the necessary data to complete the *Z*-score calculations.

Upon examination of the benchmark scores from the *PSR*, the slope appears greater for those who are not diagnosed with an SLD (Figure 2). This would suggest that those students are responding to instruction and increasing their knowledge related to reading. Further, the more flat slope for those who are diagnosed with an SLD suggests these students are not responding to the intervention provided to them. Through use of the Tukey method, the mean slope score for the *PSR* was determined to be 1.19 with a standard deviation of 0.96 and the slope coefficients score of -.003 was the equivalent of a *Z*-score of -1.27.

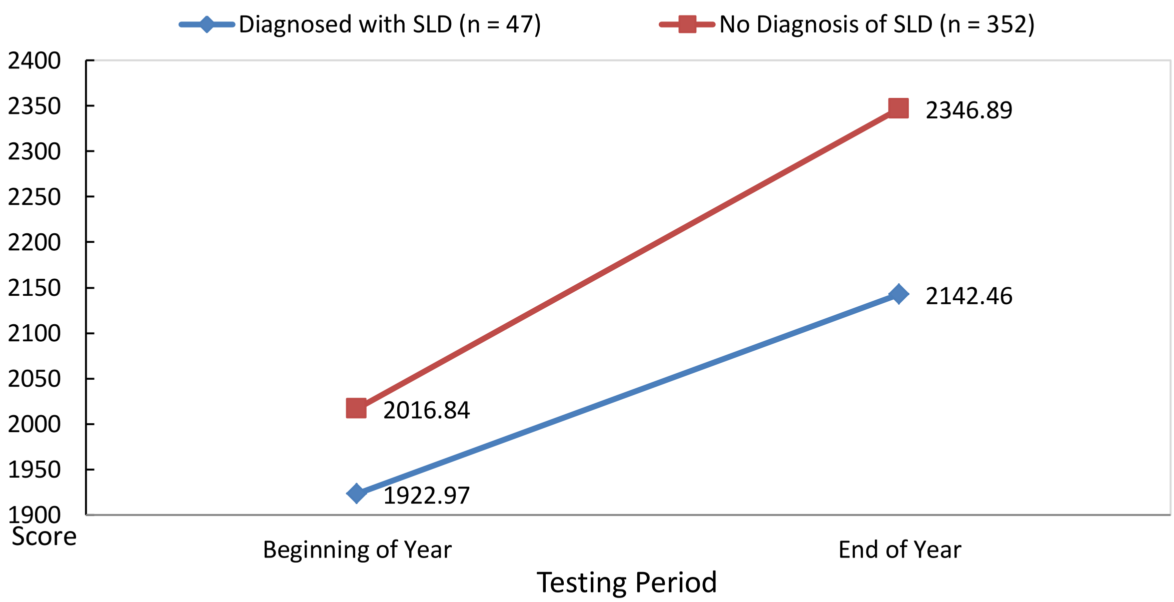

Figure 2. PSR Scores by Diagnosis.

The cut-off score for students with potential learning disabilities based on slope scores from the *PSR* and *DORF beginning of the year* and *end of the year* scores were calculated. The slope scores were converted to *Z*-scores where -1.27 was utilized as the

cut-off score (Figure 3). The mean slope score for the *PSR* was 1.19 with a standard deviation of .96 and the slope coefficients score of -.003 was the equivalent of a *Z*-score of -1.27. The *DORF* mean slope score was 0.15 with a standard deviation of 0.10 and a slope coefficient cut-off score of less than or equal to .003, which was equal to a *Z*-score of -1.27. Slope coefficient results that reflected ≤ - .003 for the *PSR* and ≤ .003 for the *DORF* were identified as the cut-off scores for determining whether a child was not making adequate progress in reading. Both of these scores reflected a flat trajectory in terms of instructional growth in reading of students with performance deficits over time. For both the *PSR* and *DORF*, students who fell below the slope cut-off scores were then classified as Slope SLD. This allowed for the comparison between the slope classification and the actual diagnosis of SLD, which was addressed in the next research question.

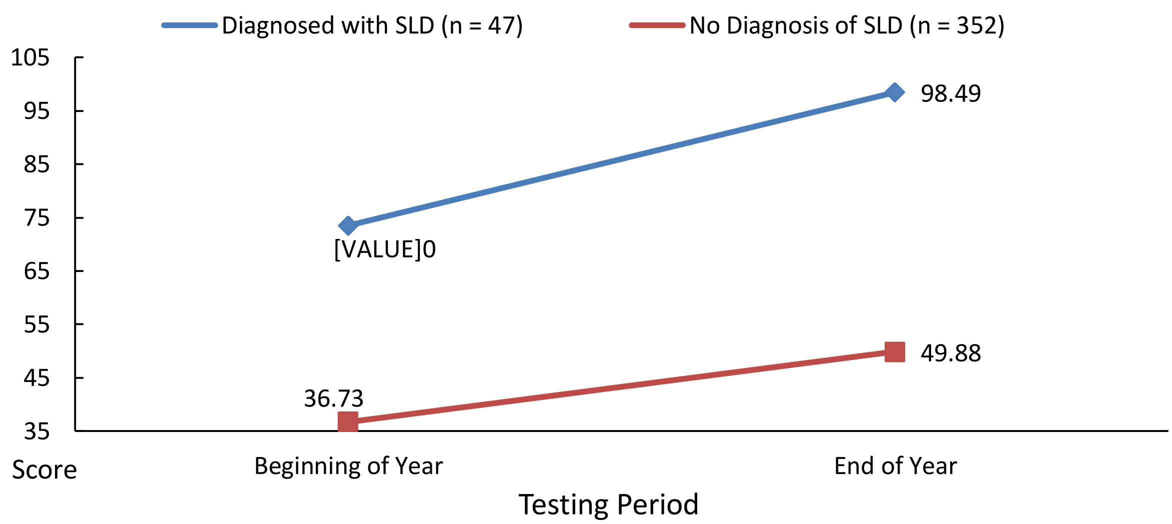

Figure 3. DORF Scores by Diagnosis.

Research Question Number Three. The last question addressed the accuracy of the combination of the *PSR* and *DORF* score discrepancy and the slope coefficient scores respectively to determine potentially improved identification of SLD. Performance results

of the locally-normed reading scores and slope differentials were used together to assess whether improved diagnostic information would result in more accurate identification of students with diagnosed SLD. This information provided another form of assessment to support IDEA 2004 legislation for the evaluation of learning disabilities. For this comparison, the student classification based on discrepancy and slope was compared to the students with an actual SLD diagnosis (i.e., Local SLD, Slope SLD, and Diagnosed SLD) through contingency correlations. For this comparison, only the *end of the year* time period was used.

There was a statistically significant relationship: $r(366) = .387$, $p < .001$ between of the *end of the year* discrepancy scores of the *DORF* measure and the students with SLD in reading (Table 11). This correlation was increased when Slope SLD was factored in ($r(350) = .552$, $p < .001$). Upon examination of slope by itself, the relationship with the diagnosis of SLD was found not to be statistically significant ($r(350) = .073$, $p < .169$).

Table 11

Correlation DORF EOY Discrepancy and Slope

Comparison	*N*	Correlation	Significance
DORF Cut-Off Score	366	0.387	0.001
DORF Slope	350	0.073	0.169
Discrepancy & Slope	350	0.552	0.001

A similar pattern was found upon examination of the relationship between the classification of SLD based on the *PSR end of the year* local classification, the *PSR* Slope SLD, and student classification based on whether or not the student was diagnosed with a

SLD (Table 13). There was a statistically significant relationship: $r(372) = .293, p < .001$ between of the *end of the year* discrepancy scores of the *DORF* measure and the students with SLD in reading. This correlation was increased when Slope SLD was factored in ($r(341) = .543, p < .001$). Upon examination of slope by itself, the relationship with the diagnosis of SLD was found to not be statistically significant ($r(341) = .050, p < .353$).

A possible reason for the lower correlation of the *PSR* slope scores was due to the student self-administration methodology of the measure. The differences of the self-administered *PSR* measures tended to show greater variance than those of individually administered tests by a trained professional. The test administrator who monitored the student test behaviors and attention to task supported the quality responses of the student. Conversely, students who took the *PSR* sat in a large room with banks of computers with children behind each of the monitors taking the test. Supervision was limited to a teacher observing the students. Any number of students tended to rush through the administration.

Another potential answer to this question proposed the use of the Rasch Model scoring system that utilized scale scores that are related to the student's grade-level learning expectations. The Rasch Model does not use standard scores that are defined by standard scores and deviation from the mean. When the *PSR* score was compared to the slope data, a statistically significant correlation was evidenced: $r(341) = .543, p < .001$. The evidence of the utilization of both forms of PM methods resulted in a moderate, but statistically significant increase in the correct classification of SLD when compared to the *PSR* discrepancy and slope scores in isolation (Table 12).

Table 12

Correlation PSR EOY Discrepancy and Slope

Comparison	*N*	Correlation	Significance
PSR Cut-Off Score	372	0.293	0.001
PSR Slope	341	0.050	0.353
Discrepancy & Slope	341	0.543	0.001

Conclusions

The current study examined a model that utilized two separate forms of reported scores by converting the results to *Z*-scores. A research-based cut-off score based on the 10^{th} percentile, -1.27 *Z*-score, was used to compare the differences between the classification of SLD between national and local norms. The reason for the comparison was to limit learning disability misidentification of students from lower socioeconomic communities with diverse social, cultural, and linguistic demographics. More specifically, local norms were compared to national norms for the identification of specific learning disabilities.

The use of local norms reduced the over-identification of SLD in the study sample. Local norms were utilized because there was the concern that national norms tended to cause educators to identify more students with learning disabilities and problematic behaviors who were exposed to lower socioeconomic environments (Utley, 2011). Historically, nationally-normed universal screening measures have a tendency to over-identify these children with potential learning problems and disabilities (Pearce & Gayle, 2009; Tran, et al., 2010). The use of local norms supported the minimization of

this problem. The first research question addressed the differences that exist in the identification of specific learning disabilities when using local norms classification compared to national norms classification.

Third-grade students in the current study who were administered the *PSR* measures were more likely to be identified with reading difficulties when national norms were used. Based on the 362 children tested at the *beginning of the year*, 49.2 % (or 178 students) fell in the identification classification of SLD using the national norms utilizing the cut-off score of a -1.27 *Z*-score. A similar trend for the *end of the year PSR* scores indicated that 42.2% (or 157 students) of the 372 students tested were identified with scores below or equal to a -1.27 *Z*-score. Conversely, 12.9% (or 48 students) were identified on the *end of the year PSR* results using local norms. The local norms were mirrored with the 47 third-grade students who were actually diagnosed with specific learning disabilities in reading. However, both of the measures, regardless of which norms were used, were not accurate with identifying the children who were diagnosed with SLD. The scores did not reliably identify children with or without SLD. In fact, the local norms had a tendency to under-identify students by not identifying students diagnosed with SLD.

The *DORF beginning of the year* scores where $N = 380$ reflected at least 13.9 % (or 53) of the children who took the reading fluency test had met the national cut-off score norms for potential reading learning disabilities. A lower score for the local norms was evidenced at 11.6 % (or 44 students). It appeared that the national norms for the *DORF beginning of the year* were similar. There was an increase to 18.0% (or 66) of the children by the *end of the year* $N = 366$ that the national norms reflected with potential

learning disabilities compared to 11.2% (or 41 students) identified by local norms. This discrepancy accounted for 66 students identified by national norms compared with 41 students classified through the use of local norms and a cut-off score of a -1.27 *Z*-score. In this case, local norms tended to under-identify students with SLD.

The number of children who were identified using both the *PSR* and *DORF* were similar when using national norms. A slightly elevated trend was evidenced with the *DORF* local norms when compared to the *PSR* local norms. The local norms utilized the cut-off *Z*-score of -1.27, or 10th percentile, as suggested by Shinn (2007). The 10th percentile and equivalent *Z*-score of -1.27 suggested assessing for potential learning disabilities for children who fell at or below this cut-off level on universal screening measures. It was also important to consider the use of several forms of evaluation data to support the diagnosis for SLD (Ball & Christ, 2012; United States Department of Education, 2013)

The second research question addressed the slope scores that were equivalent to the discrepancy scores based on standard scores for both the *PSR* and *DORF*. The slope score was calculated using the Tukey method and the scores were converted to *Z*-scores. The score was calculated based on the child's response to academic exposure over a 170-day elapse of days. The 10th percentile cut-off score that was equivalent or less than a *Z*-score of -1.27 was identified.

The *PSR* mean slope score was 1.19 with a standard deviation of .96, which equated to a slope cut-off score of < - .003, which was equal to a *Z*-score of -1.27. The *DORF* mean slope coefficient score was .15 with a standard deviation of .10 and a slope cut-off score of < .003 equivalent to a *Z*- score of -1.27. These data were used to correlate

the discrepancy and the slope scores with the children classified with learning disabilities that were addressed in the third research question.

The slope scores provided meaningful information that addressed the rate at which a child was performing academically over time. A steady upward trend that was commensurate with the student body mean score suggested that the child was performing and growing at the expected academic rate. However, scores that were below the mean and displayed a growth trajectory that was well below average provided an indication that the child was not making adequate growth. When this trend of poor growth stagnated or leveled off, the trajectory line was a flat plane. This trajectory indicated that interventions were not providing support for the child's increased academic performance, that the child's scores were well below the mean, and that evaluation for SLD was warranted.

Regarding the final research question, the local slope scores used in conjunction with discrepant scores of the PM instruments provided an improved capability to support learning disability eligibility. Using the slope scores in isolation provided no discernable predictability of a diagnosis of SLD. While the discrepant scores did show a statistically significant relationship with the diagnosis of SLD, the addition of the slope information produced a stronger relationship (i.e., higher correlation, significant level). This synergistic relationship between the use of discrepancy and slope results suggested both scores should be used to increase the accuracy of classification of SLD. To actually compare these forms of scores provided the assessor with information that was often reviewed, but seldom compared in a statistically meaningful approach. Further, while the combined effect of using both discrepancy and slope were found for both measures, the relationship was found stronger with the *DORF*.

The *DORF end of the year* performance discrepancy relationship did have a moderate relationship of: $r(366) = .387$, $p < .001$. Conversely, the *DORF end of the year* slope relationship to children with SLD reflected a weak-to-nonexistent relationship of: $r(350) = .073$, $p < .169$. Combining the discrepancy and slope scores were significant: $r(350) = .552$, $p < .001$. The moderate relationship of the *DORF* scores and slope would suggest that using both of the results provided an improved justification for the identification of specific learning disabilities than when using the *DORF* measure alone.

The same process of comparing the discrepancy and slope scores of the *PSR* measure was followed. *PSR* scores showed an inverse pattern of the *DORF* scores for the discrepancy and slope. The *PSR end of the year* relationship with children with SLD showed a weak relationship: $r(372) = .293$, $p < .001$. Slope data reflected a weak to non-existent relationship when compared to the children with SLD: $r(341) = .050$, $p < .353$. However, this data showed approximately a 65% probability that the scores could have been due to chance. The *PSR* discrepancy and slope scores reflected a moderate relationship to classify children with learning disabilities using local norms, demonstrating an increase over use of discrepancy scores alone ($r(341) = .543$, $p < .001$). This assertion was supported by the *PSR* SLD classification of 32 students using the score-slope calculation and the number of students actually diagnosed in the third grade control group. There were 47 children who were already diagnosed with learning disabilities.

Using the *PSR* scores to support the diagnosis of learning disabilities should be interpreted with caution. This cautionary message is necessary due to the evidence of the

under-identification of potential SLD in reading that was identified when local norms were assessed to examine the differences of SLD identification using local norms.

Overall, the consideration of the *DORF* and *PSR* scores for discrepancy and slope indicated evidence for their use, in conjunction with each other, to support the identification of SLD. Despite the fact that moderate relationships were identified when locally-normed discrepancy and slope data were utilized concurrently, the potential for the misidentification of SLD was inherent. This misidentification of SLD was evidenced by the number of students who were not identified who had an actual diagnosis of SLD and the number of children who met the cut-off score criteria for a diagnosis of SLD of a -1.27 *Z*-score, or 10^{th} percentile, who did not have a diagnosis of SLD. This pattern was evidenced throughout the analysis of the classification of SLD in the current study. Data alone was not the answer to the age-old question of how to diagnose specific learning disabilities. However, measures used in conjunction with other evaluation tools were substantiated.

Implications and Recommendations

When conducting this research, the author was attempting to address the concern of using two universal screening measurements for the identification of SLD in reading. The goal was to improve to the policies and procedures for a more consistent diagnostic process for the classification of students with reading learning disabilities utilizing benchmark testing scores. Based on the findings of the current study, there was no conclusive way to identify children with specific learning disabilities based on the current screening instruments utilized by the school district. Rather, guidelines that supported the decision making process for the identification of SLD need to be reexamined.

There was the consideration that the results of the research had under-identified SLD based on the third-graders who were tested. Adjusting for this condition, further research regarding cut-off scores was indicated. This consideration was evidenced on the conservative values of the standardized scores based on *Z*-score of -1.27, which was based on the 10th percentile recommendation by Shinn (2007). This was also a limitation in that the cut-off scores were set based on the 10th percentile using local norms, which previous research had shown that learning disabilities were probable (Speece, 2008).

Although there was moderate evidence based on the correlational study of the slope and discrepancy performance, caution should be exercised when focused totally on data procured through benchmark testing (Glover & Albers, 2007). Other forms of assessment can provide useful information to support the students with learning disabilities. This information includes cognitive and neurological measures (Fletcher et al., 2011). Other forms of evaluation involve the use of achievement correlates of global functioning that are based on norm-based academic and intellectual testing subscales referred to as the Cattell-Horn-Carroll Theory (Floyd et al., 2007). The assessment information that is procured through these battery approach methods provides useful information to support diagnosis of SLD. These methods are supported with extensive research and time-tested evaluation techniques.

The accepted practice of diagnosing SLD are based on benchmark and PM testing. These forms of assessment data provide information to increase or decrease the amount of intervention and to avoid the misidentification of students with potential learning disabilities in reading (Polcyn et al., 2014). However, when students are undergoing academic interventions, weekly to biweekly progress monitoring is

recommended (Ardoin & Christ, 2009; Briggs, 2011; Christ et al., 2013). Additionally, this data can help to support schools that may be struggling academically to identify where interventions within the curriculum needed to be initiated or further augmented (Deno et at., 2009).

The next consideration was the necessity for the use of local norms in order not to over-identify children with learning disabilities. The position that local norms were considered instead of national norms would indicate a potential problem with the curriculum and/or instruction that was provided to the students at the Tier-1 level of MTSS. If students were not learning at the rate expected by the national standards, despite the fact that these children come from lower socioeconomic and diverse communities, serious consideration was required to address the needs that the children possessed to meet the challenges of learning at the rate of their peers across the nation.

The use of *Z*-scores in the diagnosis of SLD was limited. When these evaluations follow the children to schools outside of the school district, there are problems with the accepting school district or school understanding the procedures and values that were used to determine SLD. It was important to add information that other school psychologists could interpret. This would include percentile scores that the psychometricians would understand and specifications of the cut-off scores that were based on local norms.

The recommendations based on the research conducted in the current study consist of continuing with the use of local norms in order to limit the over-identification of learning disabilities with the children in the school district population. Instead of addressing several scales at once, focusing on one form of measure to better understand

the evaluation characteristics of the sample was indicated. These evaluations include, not only benchmark and PM testing, but also normed-reference testing to include intellectual/cognitive functioning and achievement scales. A battery approach has provided school psychologists with a method to evaluate, not just the academic performance deficits, but also the students' global functioning. Offering this evaluation information was not just identifying a child with a disability, but also, providing a wealth of information to support the learning styles and academic needs of the student.

Another consideration involves the identification needs of the general education instruction programs and instruction. The use of local norms to minimize SLD diagnosis suggest a limitation of the school district to teach the learning standards successfully based on their curriculum. However, local norms show a favorable evaluation process of curricular needs that support changes in the curricula. The learning styles and deficits that the children present with that limit their comparability to nationally-driven norms of the screening instruments used will need further investigation. This position supports the continued use of local norms as a means to quantify progress and develop data-driven goals for the school district.

Understanding these recommendations is a process of initiating the steps to address the instructional needs of the children at the Tier-1 level. Once this developmental process improves the universal screening scores of the children, local norms should be re-assessed and possibly discontinued. During this time of improvement, the school district should use this inferential data to support and guide instruction and intervention intensity within the multi-tiered systems of support.

Overall, the current study has led to further questions and concerns about the utilization of normative instruments for the identification of SLD. Although local norms have reflected a plausible tool to minimize the over-identification of SLD in school districts with children impacted by poverty, cultural diversity, and associated environmental conditions, other concerns regarding the instructional influence of the Tier 1 supports were put into question. Regardless of the population that a school or school district serves, the instructional needs of the students are paramount. Children with learning disabilities represent only a small portion of the school population and should be evaluated with sensitivity to the students' social, economic, ethnic, and language diversity. While the use of local norms also provide some misidentification of SLD, use of local comparisons remains functional when examining students who are educated within a diverse school district. Local norms support a reduction in the over-identification of SLD, and can be used to support instructional guidance to meet the needs of the diverse instructional requisites of students.

REFERENCES

Aaron, P. G., Joshi, R. M., Gooden, R., & Bentum, K. E. (2008). Diagnosis and treatment of reading disabilities based on component model of reading: An alternative to the discrepancy model of LD. *Journal of Learning Disabilities*, *41*(1), 67-84. doi:10.1177/0022219407340838

Ahearn, E. I. (2009). State eligibility requirements for specific learning disabilities. *Communication Disorders Quarterly, 30*(2), 120-128. doi:10.1177/1525740108325221

Algozzine, B., & Ysseldyke, J., (1992). *Strategies and tactics for effective instruction.* Longmont, CO: Sopris West.

Algozzine, B., Ysseldyke, J., & Elliott, J. (1997). *Strategies and tactics for effective instruction.* (2nd ed.). Longmont, CO: Sopris West.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* [Text Revision] *4th ed*. Washington, DC: American Psychiatric Association.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders 5th*. Arlington, VA: American Psychiatric Association.

Appelbaum, M. (2009). *The one-step guide to implementing RTI: Academic and behavioral interventions, K-12*. Thousand Oaks, CA: Corwin Press.

Ardoin, S. P. (2006). The response in response to intervention: Evaluating the utility of assessing maintenance of intervention effects. *Psychology in the Schools, 43*(6), 713-725. doi:10.1002/Pitts.20181

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*(2), 266-283.

Ball, C. R., & Christ, T. J. (2012). Supporting valid decision making: Uses and misuses of assessment data within the context of RTI. *Psychology in the Schools, 49*(3), 231-244. doi:10.1002/pits.21592

Bamonto-Graney, S., Martinez, R. S., Missall, K. N., & Aricak, O. T. (2010). Universal screening of reading in late elementary school: R-CBM versus CBM maze. *Remedial and Special Education, 31*, 368-377. doi:10 .1177/0741932509338371

Batsche, G., Curtis, J., Dorman, C., Castillo, J., & Porter, L. J. (2008). The Florida problem-solving/response to intervention model: Implementing the statewide initiative. In S. Jimerson, M. Burns, & A. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 378-395), NY: Springer Publishing.

Berkeley, S., Bender, W. N., Peaster, L. G., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities, 42*(1), 85-95.

Bocian, K. M., Beebe, M. E., MacMillan, D. L., & Gresham, F. M. (1999). Competing paradigms learning disabilities classification high schools and the variations in the meaning of discrepant achievement. *Learning Disabilities Research & Practice, 14*(1), 1-14.

Bond, T. G., & Fox, C. M. (2012). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New York, NY: Taylor & Francis.

Bramlett, R., Cates, G. L., Savina, E., & Lauinger, B. (2010). Assessing effectiveness and efficiency of academic interventions in school psychology journals: 1995-2005. *Psychology in the Schools, 47*(2), 114-125. doi:10.1002/pitts.20457

Briggs, R. N. (2011). *Investigating variability in student performance on DIBELS oral reading fluency third-grade progress monitoring probes: Possible contributing factors.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 346-6319)

Brown-Childsey, R., & Steege, M. W. (2010). *Response to intervention*. New York, NY: The Guildford Press.

Burns, M. K. (2007). Reading at the instructional level with children identified as learning disabled: Potential implications for response-to-intervention. *School Psychology Quarterly, 22*(3), 297-313. doi:10.1037/1045.3830.22.3.297

Callinan, S., Cunningham, E., & Theiler, S. (2013). Revisiting discrepancy theory in learning disabilities: What went wrong and why we should go back. *Australian Journal of Guidance and Counseling, 23*(1), 1-17. doi:10.1017/jgc.2012.22

Canter, A. (1995). Best practices in developing local norms and behavioral assessment. In A. Thomas , & J. Grimes, (Eds.), *Best practices in school psychology-III* (pp. 689-700). Washington, DC: National Association of School Psychologists.

Carroll, J. (1963). A model of school learning. *Teachers College Record, 64*, 723-733.

Cavendish, W. (2013). Identification of learning disabilities: Implications of proposed DSM-5 criteria for school-based assessment. *Journal of Learning Disabilities, 46*(1), 52-57. doi:10.1177/0022219412464352

Centers for Disease Control and Prevention (2011). *Developmental disabilities increasing in US*. Retrieved from www.cdc.gov/features/dsdev_disabilities/

Chin, E., & Wong, P. (2013). Preparing teachers: Highly qualified to do what? Editors' introduction. *Education Policy Analysis Archives*, *21*(54). Retrieved from http://epaa.asu.edu/ojs/article/view/1406

Christianson, S. L., & Ysseldyke, J. E. (1989). Assessing student performance . . . An important change needed. *Journal of School Psychology*, *27*, 409-426.

Christ T. J., & Coolong-Chaffin, M. (2007). Interpretations of curriculum-based measurement outcomes: Standard error and confidence intervals. *School Psychology Forum, 1*, 75-86.

Christ, T. J., & Silberglitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review, 36*(1), 130-146.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurements of oral reading: An evaluation of growth rates and seasonal effects among students served in the general and special education. *School Psychology Review, 39*(3), 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. *Council for Exceptional Children, 78*(3), 356-373.

Christ, T. J., Zopluoglu, C., Monaghen, B. D., & Van Norman, E. R. (2013). Curriculum-based measurement of oral reading: Multi-study evaluation of schedule, duration, and data set quality on progress monitoring outcomes. *Journal of School Psychology, 51*, 19-57. doi:10.6/J.JSP.2012.11.001

Chun, M., & Witt, J. (2008). System to enhance educational performance. Retrieved from http://www.isteep.com/index.html

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., … Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage Katie screening process. *Journal of Educational Psychology, 102*(2), 327-340. doi:10.1037/a0018448

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in the first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*(2), 394-409. doi:10.1037/0022-0663.98.2.394

Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., … Bouton, B. (2012). Accelerating chronically unresponsive children to tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities*, *45*(3), 204-216. doi:10.1177/0022219412442151

Cummings, K. D., Kennedy, P. C., Otterstedt, J., Baker, S. K., & Kame'enui, E. J. (2011). *DIBELS Data System: 2010-2011 Percentile ranks for DIBELS next benchmark assessments* (Technical Report 1101). Eugene, OR: University of Oregon.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1), 1-11.

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

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., & Chaing, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*(1), 36-45.

Deno, S. L., Reschly, A. L., Lemke, E. S., Magnusson, D., Callender, S. A., Windram, H., & Satchel, N. (2009). Developing a school-wide progress-monitoring system. *Psychology in Schools, 46*(1), 44-55. doi:10.1002/pits20353

Dombrowski, S. C., Kamphaus, R. W., Barry, M., Brueggeman, A., Cavanaugh, S., Devine, K., & Vess, S. (2006). The Solomon effect in learning disabilities diagnosis: Can we learn from history? *School Psychology Quarterly*, *21*(4), 359-373.

Dombrowski, S. C., Reynolds, C. R., & Kamphaus, R. W. (2004). After the demise of the discrepancy: Proposed learning disabilities diagnostic criteria. *Professional Psychology: Research and Practice, 35*(4), 364-372. doi:10.1037/0735-7028.35.4.364

Epps, S., Ysseldyke, J., & McGue, M. (1984). Differentiating LD and non-LD students: “I know one when I see one." *Learning Disability Quarterly, 7*, 89-101.

Fagan, T. K. (1995). Trends in history school psychology United States. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology-III* (pp. 59-67). Washington, DC: National Association of School Psychologists.

Faggella-Luby, M., & Wardwell, M. (2011). RTI in the middle school: Findings and practical implications of a reading comprehension study. *Learning Disability Quarterly*, *34*(1), 35-49.

Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia*, *54*, (2), 304-331.

Fletcher, J. M., Steubing, K. K., Barth, A. E., Denton, C. A., Cirino, P. T., Francis, D. J., & Vaughn, S. (2011). Cognitive correlates of inadequate response to reading intervention. *School Psychology Review, 40*(1), 3-22.

Floyd, R. G., Keith, T. C., Tuab, G. E., & McGrew, K. S. (2007). Cattell-Horn-Carroll cognitive abilities and their effects on reading decoding skills: *g* has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly, 22*(2), 200-233. doi:10.1037/1045-3830.22.2.200

Forsyth, R., Colver, A., Alvanides, S. Wooly, M., & Lowe, M. (2007). Participation of young severely disabled children is influenced by their intrinsic impairments and environment. *Developmental Medicine and Child Neurology, 49*, 345-349.

Francis, J. D., Fletcher, J. M., Steubing, K. K., & Lyon, R. (2005). Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38,* 98-108.

Fuchs, D., & Fuchs, L. S. (2005). Responsiveness-to-intervention: A blueprint for practitioners, policymakers, and parents. *Teaching Exceptional Children, 38*, 57-61.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 92-99.

Fuchs, L. S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children, 39*(5), 14-20.

Fuchs, L. S., Fuchs, D., & Hollenbeck, K. N. (2007). Extending responsiveness to intervention to mathematics at first and third grades. *Learning Disabilities Research & Practice*, *22*(1), 13-24.

Gay, L. R., Mills, G. E., & Airasian, P. (2012). *Educational research: Competencies for analysis and applications* (10th ed.). Boston: Pearson.

Germann, G., & Tindal, G. (1985). An application of curriculum-based measurement: The use of direct and repeated measurement. *Exceptional Children, 52*, 244-265.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in primary grades. A practice guide (NCEE 2009-4045).* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Educational Sciences, U.S. Department Of Education. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practice_guides/rti_reading_pg_021809.pdf

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of school psychology, 45*, 117-135.

Glover, T. A., & Diperna, J. C. (2007). Service delivery for response to intervention: Core components and directions for future research. *School Psychology Review, 36*(4), 526-540.

Good, R. H., Gruba, J., & Kaminski, R. A. (2001). Best practices using dynamic indicators of basic early literacy skills (DIBELS) and then outcomes-driven model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-IV* (pp. 679-700). Washington, DC: National Association of School Psychologists.

Good, R., Kaminski, R., Shinn, M., Bratten, J., Shinn, M., Laimon, D., … Flindt, N. (2004). *Technical adequacy of DIBELS: Results of the early childhood research institute on measuring growth and development (Technical Report, No. 7).* Eugene, OR: University of Oregon.

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

Gresham, F. M. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review, 38*(4), 533-540.

Hadedank, L. (1995). Best practices in developing local norms for problem solving in schools. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology-III* (pp. 701-715). Washington, DC: National Association of School Psychologists.

Hagan-Burke, S., & Jefferson, G. L. (2002). Using data to promote academic benefit for included students with mild disabilities. *Preventing School Failure, 46*(3), 112-118.

Hagermoser-Sanetti, L. M., Gritter, K. L., & Dobey, L. M. (2011). Treatment integrity of interventions with children in the school psychology literature from 1995 to 2008. *School Psychology Review, 40*(1), 72-84.

Hale, J., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., . . . Yalof, J. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification intervention: An expert white paper consensus. *Learning Disability Quarterly, 33*(3), 223-236.

Harry, B., & Klinger, J. (2006). *Why are so many minority students in special education? Understanding race and disability in schools.* New York, NY: Teachers College Press.

Harry, B., & Klinger, J. (2007). Discarding the deficit model. *Educational Leadership, 64*(5), 16-21.

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

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review, 34*(3), 372-386.

Hixson, M., Christ, T. J., & Bradley-Johnson, S. (2008). Best practices in the analysis of progress-monitoring data and decision making. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-V*, (pp. 2133-2146). Bethesda, MD: NASP.

Hosp, J. L., Hosp, M. A., & Dole, J. K. (2011). Potential bias in projective validity of universal screening measures across disaggregation subgroups. *School Psychology Review, 40*(1), 108-131.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: a practical guide to curriculum-based measurement*. New York, NY: The Guilford Press.

Illinois State Board of Education (2009). *2008-2009 Annual state report on special education.* Retrieved from www.isbe.net/spec-ed/pdf_reports/annualperformance_report09.pdf

Illinois State Board of Education (2012a). *2012 annual report*. Retrieved from http://www.isbe.net/reports/annual12/special_ed.pdf

Illinois State Board of Education (2012b). *Overview of Illinois' process for determining and addressing disproportionality in special education identification and significant disproportionality in identification, placement, and discipline.* Retrieved from http://www.isbe.net/spec-ed/pdfs/disproportionality_process.pdf

Jenkins, J. R., Graff, J. J., & Miglioretti, D. L. (2009). Estimating growth using intermittent CBM progress monitoring. *Council for Exceptional Children, 75*(2), 151-163.

Jenkins, J. R., Schiller, E., Blackorby, J., Thayer, S. K., & Tilly, W. D. (2013). Responsiveness to intervention in reading: Architecture and practices. *Learning Disability Quarterly, 36*(1), 36-46. doi:10.1177/0731948712464963

Jenkins, J., & Terjeson, K. J. (2011). Monitoring reading growth: Goal setting, measurement frequency, and methods of evaluation. *Learning Disabilities Research & Practice, 26*(1), 28-35.

Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice, 24*(4), 174-185.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 59*, 421-432.

Kane, M. T. (2001). Concerns and validity theory. *Journal of Educational Measurement, 38*, 319-342.

Karelitz, T. M., Fields, E., Levy, A. J., Martinez-Gudapakkam, A., & Jablonski, E. (2011). No teacher left unqualified: How teachers and principals respond to highly qualified mandate. *Science Educator, 20*(1), 1-11.

Kavale, K. A. (2005). Identifying specific learning disability: Is responsiveness to intervention the answer? *Journal of Learning Disabilities, 38*(6), 553-562.

Kavale, K. A., & Spaulding, L. S. (2008). Is response to intervention good policy for specific learning disability? *Learning Disabilities Research and Practice, 23*(4), 169-179.

Kim, Y., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology, 102*(3), 652-667. doi:10.1037/a0019643

Klinger, J. K., & Edwards, P. (2006). Cultural considerations with response-to-intervention models. *Reading Research Quarterly, 41*, 108-117.

Kuntz, D., & Eulass, A. (2009). Educational rights and responsibilities: Understanding special education in Illinois. *Illinois State Board of Education.* Retrieved from: www.isbe.net/Speced/pdfs/parent guide_english.pdf

Lee, P. C., & Stewart, D. P. (2013). Does a socio-ecological school model promote resilience in primary schools? *Journal of School Health, 83*(11), 795-804.

Marston, D., & Tindal, G. (1995). Best practices and performance monitoring. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology-III* (pp. 597-607). Washington, DC: National Association of School Psychologists.

Marston, D., & Magnusson, D. (1988). Curriculum-based measurement: District level implementation. In J. L. Graden, J. E., Zins, M. J., & Curtis, J. (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 137-173). Washington, DC: National Association of School Psychologists.

McNamara, J. K., & Willoughby, T. (2010). A longitudinal study of risk-taking behavior in adolescents with learning disabilities. *Learning Disabilities Research & Practice, 25*(1), 11-24. doi:10.1111/j.1540-5826.2009.00297.x

Mellard, D., & Johnson, E. (2007). *RTI: A practitioner's guide to implementing response to intervention.* Thousand Oaks, CA: Corwin Press.

Mellard, D. M., McKnight, M., & Jordan, J (2010). RTI tier structures and instructional intensity. *Learning Disabilities Research & Practice*, *25*, 217-225.

Mellard, D. M., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research & Practice, 24*(4), 186-195.

Messick, S. (1989). Validation. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). National Council on Measurement in Education, American Council on Education. New York, NY: Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741-749.

Meta Metrics (2015). *The lexile framework for reading: Matching readers with texts.* Retrieved from https://www.lexile.com/about-lexile/lexile-overview/

National Center for Response to Intervention (2009). Progress monitoring: Frequently ask questions. [Electronic Version]. Retrieved from http://www.rti4success.org/index.php?option=_content&task=view&id=1010&Itemid=161

Nese, J. F. T., Park, B. J., Alonso, J., & Tindal, G. (2011). Applied curriculum-based measurements as a predictor of high-stakes assessment. *The Elementary School Journal*, *111*(4), 608-624.

No Child Left Behind Act [NCLB]. (2001). Public Law 107-15.

O'Connor, R. E., & Klingner, J. K. (2010). Poor responders in responsiveness to intervention models. *Theory Into Practice, 49,* 297-304.

Parker, R. I., Vannest, K. J., Davis, J. L., & Clemens, N. H. (2012). Defensible progress monitoring data for medium- and high-stakes decisions. *Journal of Special Education, 46*(3), 141-151. doi:10.1177/0022466910376837

Pearce, L. R., & Gayle, R. (2009). Oral reading fluency as a predictor of reading comprehension with American Indian and White elementary students. *School Psychology Review*, *38*(3), 419-427.

Petscher, Y., Cummings, K. D., Biancarosa, G., & Fien, H. (2013). Advanced (measurement) applications of curriculum-based measurement in reading. *Assessment for Effective Intervention, 38*(2), 71-75. doi:10.1 177/1534508412461434

Polcyn, D. M., Levine-Donnnerstein, D., Perfect, M. M., & Obrzut, J. D. (2014). Reading intervention and special education referrals. *School Psychology Forum: Research in Practice, 8*(3), 156-167.

Prasse, D. P., & Reschly, D. J. (1986). Larry P.: A case of segregation, testing, or program efficacy? *Exceptional Children, 52*(4), 333-346.

Reschly, D. J., Kicklighter, R., & McKee, P. (1988). Recent placement litigation: III, analysis of differences. In Larry P. Marshall and S-1 and implications for future practices. *School Psychology Review, 17*(1), 1988, 39-50.

Reschly, D. J., & Ysseldyke, J. E. (1995). School psychology paradigm shift. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology-III* (pp. 17-31). Washington, DC: National Association of School Psychologists.

Reynolds, C. R. (1992). Two key concepts in the diagnosis of learning disabilities and the habilitation of learning. *Learning Disability Quarterly, 15,* 2-12.

Reynolds, C. R. (2008). RTI, neuroscience, and sense: Chaos in the diagnosis and treatment of learning disabilities. In E. Fletcher-Jansen & C. R. Reynolds (Eds.), *Neuropsychological perspectives on learning disabilities in the area of RTI* (pp. 82-98). Hoboken, NY: John Wiley & Sons.

Reynolds, C. R., & Shaywitz, S. E. (2009). Response to intervention: Prevention and remediation, perhaps. Diagnosis, no. *Child Development Perspectives, 3*(1), 44-47.

Roberts, G., Good, R., & Corcoran, S. (2005). Story retell: A fluency-based indicators of reading comprehension. *School Psychology Quarterly, 20,* 304-317.

Rosenberg, L., Bart, O., & Ratzon, N. Z. (2013). Personal and environmental factors predict participation of children with and without mild developmental disabilities. *Journal of Child and Family Studies*, *22*, 658-671. doi:10.1 007/s10826-012-9619-8

Roehrig, A. D., Duggar, S. W., Moats, L., Glover, M., & Mincey, B. (2008). When teachers work to use progress monitoring data to inform literacy instruction identifying potential supports challenges. *Remedial and Special Education, 29*(6), 364-382. doi:10.1177/0741932507314021

Salkind, N. J. (2012). *Exploring research* (8th ed.), Boston, MA: Pearson.

Salvia, J., & Ysseldyke, J. P. (1998). *Assessment* (7th ed.), NY: Houghton Mifflin.

Sandoval, J. (1993). The history of interventions in school psychology. *Journal of School Psychology, 31*, 195-217.

Scanlon, D. (2013). Specific learning disability in its newest definition. Which is comprehensive? And which is insufficient? *Journal of Learning Disabilities. 46*(26), 26-33. doi:10.1177/0022219412464342

Scott, D. G., & Weishaar, M. K. (2003). Curriculum-based measurement for reading progress. *Intervention in School and Clinic, 38*(3), 153-159.

Shapiro, E. S., Hilt-Panahon, A., Gischlar, K. L., Semenniak, K., Leichman, E., & Bowles, S. (2011). An analysis of consistency between team decisions and reading assessment data within an RTI model. *Remedial and Special Education, 33*(6), 335-347. doi:10.1177/0741932510397763

Shaw, R., & Shaw, D. (2002). Technical report DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State assessment program (CSAP Technical Report). Eugene, OR: University of Oregon.

Shinn. M., & Marston, D. (1985). Using curriculum-based measures to identify mildly handicapped students. *Remedial and Special Education, 6*(2), 31-45.

Shinn, M. (1988). Development of curriculum-based norms for use in special education decision-making. *School Psychology Review, 17*, 61-80.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children.* New York, NY: Guilford Press.

Shinn, M. R. (2001). Best practices in using curriculum-based measurement and problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-IV* (pp. 671-697). Washington, DC: National Association of School Psychologist.

Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within response to intervention: Research on educational need and benefit from academic intervention. *School Psychology Review, 36*(4), 601-614.

Sibley, D., Biwer, D., & Hesch, A. (2001). Establishing curriculum-based measurement oral reading fluency performance standards to predict success on local and state tests of reading achievement. *Unpublished Data.* Arlington Heights, IL: AHSD 25.

Skiba, R. J., Poloni-Staudinger, L., Simmons, A. V., Feggins-Azziz, R., & Chung, C. G. (2005). Unproven links: Can hardly explain ethnic disproportionality special education? *Journal of Special Education, 39*, 130-144.

Speece, D. L. (2008). Learning disabilities and the USA: Operationalizing a construct. In L. Florian & M. J. McLaughlin (Eds.), *Disability classification in education: Issues and perspectives* (pp. 69-82). Thousand Oaks, CA: Corwin Press.

Stecker, P., Lemke, E. S., & Foegan, A., (2008). Using progress-monitoring data to improve instructional decision-making. *Preventing School Failure, 52*(2), 48-58.

Sternberg, R. J. (1983). Criteria for intellectual skills training. *Educational Researcher, 12*, 6-12.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* NY: Cambridge University Press.

Steubing, K. K., Fletcher, J. M., LeDoux, J. M., Reid Lyon, G., Shaywitz, S. E., & Shaywitz, B. A. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal, 39*, 469. doi:10.3102/00028312039002469

Stewart, L. H., & Silberglitt, B. (2008). Best practices in developing academic local norming. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-V* (pp. 225-242). Bethesda, MD: NASP.

Stollar, S. A., Schaeffer, K. R., Skelton, S. M., Stine, K. C., Lateer-Huhn, A., & Poth, R. L. (2008). Best practices in professional development: An integrated, three-tier model of academic and behavioral supports. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-V* (pp. 875-886). Bethesda, MD: NASP.

Stoodley, C. J., Ray, N. J., Jack, A., & Stein, J. F. (2008). Implicit learning in control, dyslexic, and garden-variety poor readers. *Annals of the New York Academy of Sciences, 1145*, 173-183.

Sullivan, A. L., & Artiles, A. J. (2011). Theorizing racial inequity in special education: Applying structural inequity theory to disproportionality. *Urban Education*, *46*, 1526-1552. doi:0042085911416014.

Sullivan, A. L., A'Vant, P., Baker, J., Chandler, D., Graves, S., McKinney, & Sayles, T. (2009). Confronting inequity in special education, Part I: Understanding the problem of disproportionality. *Communiqué, 38*(1). Retrieved from http://www.nasponline.org/publications/cq/38/1/disproportionality.aspx

Swanson, E., Solis, M., Ciullo, S., & McKenna, J. W. (2012). Special education teachers' perceptions and instructional practices in response to intervention implementation. *Learning Disability Quarterly, 35*(2), 115-126. doi:10:1177/0731948711432510

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1,* 1-26.

Tannock, R. (2013). Rethinking ADHD and LD in DSM-5: Proposed changes in diagnostic criteria. *Journal of Learning Disabilities, 46*(5), 5-25. doi:10.1177/0022219412464341

Therrien, W. J., & Hughs, C. (2008). Comparison of repeated reading and question generation on students reading fluency and comprehension. *Learning Disabilities: A Contemporary Journal, 6*(1), 1-16.

Tichà, R., Espin, C. A., & Wayman, M. M. (2009). Reading progress monitoring for secondary-school students: Reliability, validity, and sensitivity to growth of reading-aloud and maze-selection measures. *Learning Disabilities Research & Practice*, *24*(3), 132-142.

Tran, L., Sanchez, T., Arellano, B., & Swanson, H. L. (2010). A meta-analysis of the RTI literature for children at risk for reading disabilities. *Journal of Learning Disabilities, 44*(3), 283-295. doi:10.1177/0022219410378447

Tukey, J. W. (1977). Straightening out plots (using three points). In Exploratory data analysis (pp. 169-181). Menlo Park, CA: Addison-Wesley.

Tyler, N. C., Yzquierdo, Z., Lopez-Reyna, N., & Flippin, S. S. (2004). Cultural and linguistic diversity and special education workforce: A critical overview. *Journal of Special Education,* 38(1), 22-38.

United States Census Bureau (2012). Income and poverty in the United States 1966 - 2012. Retrieved from: http://www.census.gov/prod/2014pubs/p60-248.pdf

United States Department of Education. (2013). *Individuals with disabilities education improvement act* (2004). Retrieved from http://idea.ed.gov/explore/home

University of Oregon Center on Teaching and Learning (2012). *DIBELS next recommended benchmark goals: Technical supplement* (Technical Report 1204). Eugene, OR: Author.

Utley, C. A. (2011). The psychometric investigation of multicultural and special education survey: An exploratory factor analysis. *Learning Disabilities: A Contemporary Journal, 9*(1), 47-70.

Utley, C. A., Oiakor, F. E., & Bakkon, J. P. (2011). Culturally responsive practices for culturally and linguistically diverse students with learning disabilities. *School Psychology Review, 36*(4), 541-561.

Van Der Heyden, A. M. (2011). Evolving standards of diagnostic accuracy in predicting and avoiding academic failure. *Advances in Learning and Behavioral Disabilities, 24,* 59-78. doi:10.1108/S0735-004X(2011)0000024005

Villegas, A. M., & Irvine, J. J. (2010). Diversifying the teaching force: An examination arguments. *Urban Review, 42,* 175-192.

Wasnek, J., & Vaughn, S. (2007). Research-based implications from extensive early interventions. *School Psychology Review, 36*(4), 541-561.

Wedl, J. R. (2005). *Response to intervention: An alternative to traditional eligibility criteria for students with disabilities*. Minneapolis, MN: Center for Policy Studies and Hamline University.

Yeo, S., Fearington, J., & Christ, T. J. (2011). An investigation of gender, income, and special education status bias on curriculum-based measurement slope in reading. *School Psychology Quarterly, 26*(2), 119-130. doi:10.1037/a0023021

Ysseldyke, J., Burns, M. K., Scholin, S. E., & Parker, D. C. (2010). Structurally valid assessment within response to intervention. *TEACHING Exceptional Children, 42* (4), 54-61.

Ysseldyke, J. E., & Thurlow, M. (1984). Assessment practices in special education: Adequacy and appropriateness. *Educational Psychologist, 19*, 123-137.

Ysseldyke, J. E., Thurlow, M., Graden, J., Wesson, C., Algozzine, B., & Deno, S. (1983). Generalizations from five years of research on assessment and decision-making: The University of Minnesota Institute. *Exceptional Education Quarterly, 4*, 75-93.

Zhang, C. (2007). Response to intervention. In A. M. Bursztyn (Ed.), *The Praeger handbook of special education* (104-106). Westport, Connecticut: Greenwood.

Zirkel, P. A. (2010). The legal meaning of specific learning disability for special education eligibility. *TEACHING Exceptional Children, 42*(5), 62-67.

Zirkel, P. A. (2013). The role of the DSM in IDEA case law. *Communiqué, 39*(5). Retrieved from http://www.nasponline.org/publications/cq/39/5/RoleofDSM.aspx

Zirkel, P. A., & Rose, T. R. (2009). Scientifically based research and peer-reviewed research under IDEA. *Journal of Special Education Leadership, 22*(1), 36-50.