Integration of Traditional Assessment and Response to Intervention in Psychoeducational Evaluations of Culturally and Linguistically Diverse Students

by

Jennifer A. Planck

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2014 by the
Graduate Supervisory Committee:

Linda C. Caterino, Chair
Jill Stamm
Sylvia A. Cohen

ARIZONA STATE UNIVERSITY

December 2014

ABSTRACT

The popularity of response-to-intervention (RTI) frameworks of service delivery has increased in recent years. Scholars have speculated that RTI may be particularly relevant to the special education assessment process for culturally and linguistically diverse (CLD) students, due to its suspected utility in ruling out linguistic proficiency as the primary factor in learning difficulties. The present study explored how RTI and traditional assessment methods were integrated into the psychoeducational evaluation process for students suspected of having specific learning disabilities (SLD). The content of psychoeducational evaluation reports completed on students who were found eligible for special education services under the SLD category from 2009-2013 was analyzed. Two main research questions were addressed: how RTI influenced the psychoeducational evaluation process, and how this process differed for CLD and non-CLD students. Findings indicated variability in the incorporation of RTI in evaluation reports, with an increase across time in the tendency to reference the prereferral intervention process. However, actual RTI data was present in a minority of reports, with the inclusion of such data more common for reading than other academic areas, as well as more likely for elementary students than secondary students. Contrary to expectations, RTI did not play a larger role in evaluation reports for CLD students than reports for non-CLD students. Evaluations of CLD students also did not demonstrate greater variability in the use of traditional assessments, and were more likely to rely on nonverbal cognitive measures than evaluations of non-CLD students. Methods by which practitioners addressed linguistic proficiency were variable, with parent input, educational history, and

individually-administered proficiency test data commonly used. Assessment practices identified in this study are interpreted in the context of best practice recommendations.

DEDICATION

I dedicate this work to my mother, Merrily Ann Planck. Without her love and support, I wouldn't be where I am today.

ACKNOWLEDGMENTS

I would like to thank Dr. Linda Caterino, not only for her support throughout the process of completing my dissertation, but also for standing by the students of the ASU school psychology program when we needed it most. I would like to thank Dr. Sylvia Cohen for her willingness to serve on my committee, as well as the excellent supervision she has provided me with as a psychologist. Additionally, I would like to thank Dr. Jill Stamm for being a wonderful mentor throughout my graduate coursework and teaching experiences. Finally, I am eternally grateful to the family and friends who have supported me over the course of this journey. To my uncle Dennis: thank you for always believing that this day would come. And to my friends both near and far away: whether you answered my silly questions, sent me tweets of support, or took me out for a much needed mental break, your kindness and generosity have been much appreciated.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Educational Law in the United States

Free and accessible public education for children and adolescents is considered a hallmark of modern American society. Historically, however, those with special learning needs were not always accepted within our schools. It was not until the 1970s that the United States government formally introduced legislation to protect the rights of students with disabilities and ensure that they would receive quality educational experiences. Responding to expanding case law on the issue and calls from professionals in the field, President Gerald R. Ford signed the Education of All Handicapped Children Act (EAHCA, 1975). Like most American educational statutes, this act made federal education funding to states contingent on their abiding by a set of regulations. EAHCA (1975) specifically mandated that states provide a free and appropriate public education in the least restrictive setting to all students, regardless of their disability status. The law formed the basis of a special education system, still dominant today, in which students with suspected special needs are assessed by school-based multidisciplinary teams that typically include general education teachers, special education teachers, administrators, parents, and school psychologists. Notably, assessment of eligibility under EAHCA and its predecessors is two-pronged, with teams determining whether or not students are both eligible for, and in need of, special education services in order for them to benefit from the school curriculum (Yell & Drasgow, 2007).

Since its inception, EAHCA (1975) has been reauthorized and modified periodically by the U.S. government to include additional regulations governing the identification and legal rights of children with disabilities in schools. The first two reauthorizations, which took place in 1990 and 1997 respectively, led to the renaming of the law as the Individuals with Disabilities Education Act (IDEA, 1997, 1999). Provisions were created for the early identification of children with developmental delays, the protection of parental rights in school disputes, and the creation of post-high school transition programs for special needs children (U.S. Office of Special Education Programs, 2000). Additionally, the removal of the term "handicapped" from the Act's title was indicative of a greater emphasis on the needs of the individual student. Although the requirement that every child deemed eligible for special education services receive an Individualized Education Plan (IEP) had been part of the original EAHCA, the 1997 reauthorization outlined a strict set of rules and timelines for IEP creation and review (George, 1999). IEPs must outline a child's current performance, establish measurable goals, discuss the special education and related services (e.g., speech-language therapy, occupational therapy) the child will receive, outline the percentage of time the student will spend with typical peers, and be reviewed annually to determine whether or not progress is being made (IDEA, 1997).

Despite the great progress made within the first few decades after the creation of EAHCA (1975), the most recent reauthorization of the law as the Individuals with Disabilities Education Improvement Act (IDEIA, 2004) arguably carries with it the greatest implications for how schools presently identify and educate students with special

needs and, in particular, those with specific learning disabilities. In addition to aligning special education law with the No Child Left Behind Act (NCLB, 2001), IDEIA (2004) defines a child with a disability as the following:

> A child evaluated in accordance with Sec. Sec. 300.304 through 300.311 as having mental retardation, a hearing impairment (including deafness), a speech or language impairment, a visual impairment (including blindness), a serious emotional disturbance (referred to in this part as "emotional disturbance"), an orthopedic impairment, autism, traumatic brain injury, an other health impairment, a specific learning disability, deaf-blindness, or multiple disabilities, and who, by reason thereof, needs special education and related services. (20 USC § 1401(26))

Thus IDEIA (2004) establishes 13 special education categories special education services. This is, in and of itself, did not represent a great departure from the standards outlined in IDEA (1997). However, IDEIA differed from its predecessors in two fundamental ways: its exclusionary criteria for the establishment of disability status and its recommendations for the process of identifying students with specific learning disabilities (Yell & Drasgow, 2007). The former addition to the law applied to all 13 disability subgroups outlined, specifying that students could not be identified as being eligible for and in need of special education services if the primary factor in their performance, as determined by the multidisciplinary team, was a lack of appropriate instruction in reading or math, or limited proficiency in English (IDEIA, 2004). The change reflected both the increased emphasis on teacher accountability stemming from

NCLB (2001), as well as the increasing numbers of CLD students in the United States (Wagner, Francis, & Morris, 2005). In contrast, the latter addition applied specifically to students under consideration for a specific learning disability (SLD) classification. It was a reflection of the ongoing disagreement within the field regarding how to identify and treat children with learning disabilities, an area of debate that had begun before the adoption of the original EAHCA, and would continue long after the adoption of IDEIA in 2004 (Hale et al., 2010).

Specific Learning Disabilities

Of all of the areas of disability addressed in U.S. special education law, SLD has been the area marked by the greatest conflict and least consensus. Yet the very concept of learning disabilities predates, and goes beyond, its legal implications. Thus, when discussing learning disabilities, it is important to differentiate between educational classification and psychiatric diagnosis. Psychiatrically, the *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (DSM-5; American Psychiatric Association, 2013) forms the basis for classifying learning disabilities (learning disorders, by its terminology). Educationally, SLD is defined by the regulations put forth in IDEIA. The two are neither synonymous nor mutually exclusive; however, they do share a common history. Though psychiatric evaluations and diagnoses of reading, writing, and math disorders may be relevant for school-based practitioners, ultimately it is the guidelines outlined in IDEIA that determine the provision of special education services.

Early foundations. The origins of the learning disability concept go back in history far beyond the adoption of special education law. Even before having a specific

term to describe it in the literature, clinicians were documenting cases of children who, paradoxically, appeared to have reasoning capacities in the average to above average range, and yet still struggled to master basic academic skills (Fuchs, Mock, Morgan, & Young, 2003). In the 1920s, Samuel Orton became the first major American scholar to consider this issue in great depth from an empirical perspective. Orton, a physician, collected data on struggling young readers and noted specific patterns in the errors they made while attempting to decipher words, including their tendency to reverse words and letters (Hallahan & Mercer, 2002). He theorized that reading difficulty actually stemmed from faulty patterns of hemispheric dominance in the brain (Orton, 1925). In contrast to his predecessors in the field, Orton argued that simply teaching children to memorize words by sight was not enough – rather, he believed that children must first learn to associate sounds with letters and then to blend sounds together (Hallahan & Mercer, 2002). Although Orton's original term for disabilities in reading, "strephosymbolia," is no longer used today, his phonics-based approach to reading instruction left a significant legacy, and is generally accepted as best practice today (Joseph, 2008).

The first documented usage of the phrase "learning disability" itself did not come until decades after Orton's original work on the subject. The term was first utilized by Samuel Kirk, a clinical psychologist, in 1963 in a speech at a conference focused on parent advocacy (Kavale, 2002; Kirk, 1977). Drawing upon literature on brain dysfunction produced in the 1940s and 1950s, Kirk connected brain damage to unexpected learning failure in school, but argued that simply referring to all underachieving youth as "brain injured" masked important differences among them

(Kirk, 1977). He coined the term learning disability to describe those children who possessed no notable sensory handicaps or forms of mental retardation, but whose performance still indicated that they had "disorders in development in language, speech, reading, and associated communication skills needed for social interaction" (Kirk, 1963, p. 3). This conceptualization of a learning disability as a form of unexpectedly low performance in an academic domain, given a student's apparent intellectual and sensory adequacy, would eventually dominate the field and inform the design of decades of future research in this area.

The first large-scale research study of the learning disability concept was published in 1975, the same year of the passage of EAHCA. Responding to increasing questions about the existence of children whose difficulties with reading were unexpected given their apparent overall cognitive ability, Rutter and Yule (1975) undertook a study of the intellectual ability and reading skills of several hundred children from the Isle of Wight. After regressing IQ scores on the reading scores, Rutter and Yule yielded data that, contrary to expectation, were not normally distributed. Rather, the curve was distorted somewhat at the lower end, with a higher-than-anticipated number of children who were expected to perform far better in reading, given their IQ, than they actually performed. Rutter and Yule interpreted these findings as concrete, statistical support for what experts in the field had speculated – that there was a key difference between children with "general reading backwardness" (low reading skills and low cognitive ability), and those with "specific reading retardation" (low reading skills and average to high cognitive ability; Rutter & Yule, 1975, p. 181).

Learning disabilities in federal legislation: Then and now. By the time of the passage of EAHCA (1975), the dominant model for defining learning disabilities had become one of discrepancy – students with learning disabilities were those whose academic performance was significantly lower than what would be expected of them given their intellect. Although EAHCA included Specific Learning Disability (SLD) as a diagnostic category eligible for special education funding, the U.S. Office of Education (USOE) did not formally attach a definition to the statute until 1977 (Hallahan & Mercer, 2002). This definition has remained a part of federal special education law throughout its reauthorizations. Thus IDEIA (2004) mirrors the original wording nearly perfectly, and defines SLD as:

> …a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations. The term includes such conditions as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. The term does not include learning problems that are primarily the result of visual, hearing, or motor disabilities; of intellectual disability; of emotional disturbance; or of environmental, cultural, or economic disadvantage." (20 USC § 1401(30))

IDEIA (2004) also, for the most part, retains the original delineation of multiple academic areas in which SLD may occur. These were initially oral expression, listening comprehension, written expression, basic reading skill, reading comprehension,

mathematics calculation, and spelling (USOE, 1976). However, when the regulations were formally rereleased in 1977, spelling was permanently removed as a formal category of its own (USOE, 1977).Although the majority of the early research on learning disabilities focused on reading, Kirk (1962) had argued that the term learning disability could also refer to unexpectedly low performance in speech, language, writing, and mathematics. Following their inclusion in the law, these additional areas became the focus of greater study, particularly with regard to effective intervention strategies (Hallahan & Mercer, 2002).

The 2004 reauthorization of IDEIA and its associated regulations made two minor changes to the subcategories of SLD. First, mathematics reasoning was renamed mathematics problem solving (IDEIA, 2004). Second, and more importantly, IDEIA (2004) included an eligibility subcategory for reading fluency, bringing the total number of disability areas to eight. Although some argued that a deficit in reading fluency, in and of itself, did not constitute a disability, others contended that fluency data was valuable in that it could be quickly assessed and was relevant to classroom instruction (U.S. Department of Education, 2006). Moreover, reading fluency – defined as reading quickly, accurately, and expressively – had been identified as one of five crucial components of literacy instruction by the National Reading Panel (National Institute of Child Health and Human Development [NICHD], 2000).

Yet the biggest difference between the IDEIA (2004) approach to SLD and the stance adopted in earlier iterations of the law lay in the specifications regarding how the presence or absence of SLD is determined via assessment. Originally, the law reflected

the dominant perspective of the era in which it was written, and specified that school teams should determine a child had an SLD if the student had "a severe discrepancy between achievement and intellectual ability" in one of the original seven areas previously discussed (USOE, 1977, p. 65083). IDEIA (2004) changed this by specifically prohibiting states from requiring the use of a discrepancy model in SLD determination. Under the new law, although states could not compel school districts to use a discrepancy model, they could legally prohibit states from doing so; if a state chose to make no further specification, then under federal law, school districts were technically free to use any method to determine SLD eligibility, including the traditional ability-achievement discrepancy model (Yell & Drasgow, 2007).

Although IDEIA (2004) failed to set a firm standard for SLD determination practices, it did include a landmark recommendation regarding how school districts should make eligibility decisions. Citing widespread criticism of existing practices and emerging research, lawmakers specified that school districts may use "a process that determines if the child responds to scientific, research-based intervention as part of the evaluation process" (IDEIA, 2004, 20 USC § 1401(30)). Further regulations emphasized the importance of documentation if such a method were to be used, including records of the evidence-based instructional strategies employed and systematic data on student progress (U.S. Department of Education, 2006). This model of service delivery, in which students would be identified as struggling early on, exposed to appropriate evidence-based interventions, and monitored to determine the amount of progress made, had come to be known in the literature as response-to-intervention (RTI; Yell, Shriner, &

Katsiyannis, 2006). Although the concept of RTI had been discussed in the literature for many years under various names, IDEIA's (2004) endorsement of RTI as a valid alternative to a discrepancy model and a valuable component of the eligibility determination process as a whole signaled a major transition in how the American educational system conceptualized learning disabilities.

Response-to-Intervention Models

The movement toward RTI models of service delivery emerged from many sources and was in progress long before the development of IDEIA. Even before it was legally permitted for use as the sole method of identifying students with SLD, professionals in the field of school psychology were already calling for response-to-intervention data to be considered as a core component in SLD assessment protocols (Speece, Case, & Eddy Molloy, 2003). For many, RTI has been seen as a way to resolve the flaws in the focused model of SLD identification that emerged over the past few decades. Yet, although its role as an alternative to strict IQ/achievement discrepancy practices greatly contributed to its popularity, RTI became much more than that, emerging as a new way of thinking about the continuum of general and special education services that schools provide.

Critique of existing SLD definitions and practices. As RTI is so frequently juxtaposed with discrepancy-based practices, it is necessary to briefly address the flaws in those practices that have led to calls for a new model of SLD identification. These flaws begin with the way SLD has been legally defined. As it is presently written, the legal definition of a student with SLD is exclusive rather than inclusive, and defines SLD

more in terms of what it is not rather than what it actually is (Kavale, 2002). The legal subdivision of SLD into eight different categories is also problematic in that it conflicts with research findings on subtypes of reading and math disabilities (Geary, 1993; Morris et al., 1998). As Dombrowski et al. (2006) stress, existing findings on categorization are far more consistent with the DSM conceptualization of learning disabilities than the one spelled out in federal law, as they support the existence of broad types of learning disabilities in reading, writing, or math, rather than eight distinct subtypes. Moreover, the idea that students with ability/achievement discrepancies fundamentally differ from those who do not exhibit such discrepancies has been challenged, with studies showing that these two groups are more similar than different in terms of their academic achievement, performance in specific cognitive domains (e.g., phonological awareness), and capacity to respond to remediation (Steubing, Barth, Molfese, Weiss, & Fletcher, 2009; Steubing et al., 2002; Vellutino, Scanlon, & Lyon, 2000). Finally, states' and school districts' varying interpretations of the IDEIA definition have contributed to significant variability in the percentage of students identified as learning disabled – a figure that, given its theoretical nature, should not be as variable as it is (Kavale, Holdnack, & Mostert, 2005).

Criticisms of the discrepancy approach to SLD also pertain to how it is typically implemented. Vaughn and Fuchs (2003) describe this approach as a "wait to fail" model – students are not identified for special education services until they can manifest a large enough difference between how they are presently doing and how they "should" be doing. This process can go on for years, and arguably does a disservice to students, as they must fall far enough behind before action is taken to remediate their difficulties. It

has also been suggested that this model contributes to the existing disproportionality in identification of students from ethnic minority backgrounds as learning disabled (Coutinho & Oswald, 2000; Hale et al., 2010). The discrepancy model's reliance on teacher referrals and cognitive testing likely explains this relationship. For example, research has indicated that students from African American backgrounds are more likely to be referred for special education evaluations (Hosp & Reschly, 2003). Moreover, students from African American and Latino backgrounds tend to display lower performance on standardized IQ tests, the reasons for which have been highly debated (Bartholomew, 2004).

A final area in which the discrepancy model has been disputed lies in its psychometric aspects. The precise methods by which clinicians assess differences between IQ and achievement vary widely and often affect the validity of the conclusions that can be drawn (Hale et al., 2010). For example, some elect to use nonverbal scores in lieu of a full scale IQ score for students with linguistic delays, which has implications for reliability and validity. Additional psychometric issues include the reduced reliability inherent whenever difference scores are considered, as well as the phenomenon of regression to the mean (Sternberg & Grigorenko, 1999). The Flynn Effect, the documented trend toward increasing IQ scores in the population across history, also poses significant issues for learning disability assessment. Recent work has shown that not only has the Flynn Effect continued into the 21st century, but that IQ test scores, and therefore decisions of disability status, vary significantly depending on how long it has

been since a test has been renormed (Kanaya & Ceci, 2012; Skirbekk, Stonawski, Bonsang, & Staudinger, 2013).

In sum, critique of the discrepancy model for learning disability identification is multifaceted and has arisen from decades of research in the field and observations about assessment practices in schools. Even so, it has remained the dominant model of SLD assessment for reasons that go beyond its legal endorsement. In theory, use of strict IQ/achievement discrepancy formulas provides a way to standardize SLD assessment and reduce the confounding role of clinical judgment (Kavale, 2002). This approach also allows for distinguishing learning disabled students from those whose below-average academic achievement is on par with their cognitive ability (termed "slow learners" in the literature), a distinction that is important to make for both theoretical and practical purposes (Kavale et al., 2005; Wodrich & Schmitt, 2006). Yet perhaps most importantly, conceptualizing SLD in terms of cognitive and academic skills prompts practitioners to conduct comprehensive intellectual and academic assessments, which allow for a better understanding not only of whether or not a child qualifies for special services, but also of precisely why the child is unable to learn at an appropriate level and pace (Flanagan, Ortiz, & Alfonso, 2013; Hale, Kaufman, Naglieri, & Kavale, 2006). These factors, as well as the lack of a universally accepted alternative, have led to the continued prominence of the discrepancy model in SLD identification. At the same time, the significant shortcomings of this model have led to calls for a better way to identify students who are struggling and provide them with the specialized instruction they need without waiting for them to fail.

Defining RTI. Although no single, universally accepted definition for RTI currently exists, several core principles underlie the concept. The National Center on Response to Intervention (NCRTI; 2010), an initiative funded by the U.S. Department of Education for the purpose of assisting states in developing strong RTI frameworks, asserts that RTI is an instructional framework designed to help all students succeed by identifying those at risk of negative learning outcomes, providing empirically-supported interventions for them, and continually modifying the nature and intensity of those interventions based on how well students respond. In the broadest sense, RTI may be viewed as a method of transforming education into a multilevel system of prevention (Fuchs, Fuchs, & Compton, 2012). RTI has implications for every student and staff member in a school, and thus it requires a comprehensive plan of implementation and investment from principals, psychologists, special and general education teachers, and parents (Nellis, 2012). Though RTI is most commonly used in elementary schools, its use has also been documented in preschool (Bayat, Mindes, & Covitt, 2010), middle school (Dulaney, 2013; Prewett et al., 2012; Vaughn et al., 2010), and high school (Fisher & Frey, 2011) settings.

In practice, RTI typically translates to tiered levels of support, with each ascending tier describing a more intensive level of academic support and, ideally, serving a smaller number of students. For example, in the popular three-tiered model of RTI originally conceptualized by Fuchs and Fuchs (2005), tier 1 includes quality general education instruction for all students, tier 2 includes specialized intervention only provided to students deemed at risk of failure, and tier 3 refers to the small number of

students who, having not responded to previously implemented intervention, require more intensive intervention or special education services to make progress in school. The notion of a three-level model, though not universally utilized, has been adopted by many clinicians and researchers. The percentages of students receiving tier 2 and 3 levels of support are dependent on the needs of the population served, but generally fall between 15-20% and 0-5% respectively (Burns & Coolong-Chaffin, 2006; Fuchs, Fuchs et al., 2012; VanDerHeyden, Witt, & Naquin, 2003). Targeted tier 2 interventions are often delivered to students in small groups, whereas more intensive, tier 3 interventions are frequently delivered on an individual basis (Burns & Coolong-Chaffin, 2006).

RTI is often conceptualized from a problem solving perspective. This viewpoint is derived from the literature on behavioral consultation models, which describes an ongoing trial-and-error process of problem identification, problem analysis, plan implementation, efficacy evaluation, and plan modification until the problem is resolved (Deno, 2002). In RTI, professionals identify students throughout the school who are having difficulties learning and implement increasingly rigorous, empirically-based interventions to resolve their difficulties (Fuchs et al., 2003). When viewed from a problem solving standpoint, RTI is extremely individualized. As students require higher levels of support, both the intervention techniques they receive and the methods of measuring their progress become more closely matched to the student's own strengths and weaknesses (Hale et al., 2006). It is assumed that if students do not respond to extremely individualized modifications in the general education setting, consideration of special education services is warranted (Fuchs, Fuchs, & Compton, 2004).

An alternative way of conceptualizing RTI is known as the standard protocol method. Like advocates of the problem solving perspective, proponents of the standard protocol approach to RTI believe strongly in identifying struggling learners, implementing empirically-based interventions and continually monitoring progress (Vaughn, Linan-Thompson, & Hickman, 2003). However, they argue that it is much more efficient for all children with the same type of academic problem to receive the same intervention, rather than a highly individualized intervention package (Fuchs et al., 2003). The emphasis is on uniformity and standardization – students are continually assessed with the same curriculum-based measures, their progress is compared with normative data, and judgments of progress are made based on growth curve models (Hale et al., 2006). Use of a standard treatment protocol (including standardized interventions and a streamlined general curriculum) for all students is thought to be the best way of ensuring that inadequate instruction is not to blame for nonresponsiveness (Fuchs et al., 2004). Moreover, when intervention protocols are more standardized, teachers should require less training, and it is easier to determine whether the overall RTI system is effective (Fuchs et al., 2003).

RTI implementation: Themes and variations. Just as there is no standard definition of RTI, there is also no single model for how it should be implemented, and school systems vary widely in how they define and practice RTI. Many differences are simply in semantics. For example, while the three-tiered model is the most common conceptualization of an RTI system, the levels themselves may be referred to as primary, secondary, and tertiary prevention (Fuchs, Fuchs et al., 2012), or as universal, targeted,

and intensive intervention (Burns & Coolong-Chaffin, 2006). Yet other differences are more significant, and have become points of lingering debate and extensive research within the field.

Fuchs et al. (2003) produced a list of five steps central to any RTI scheme, regardless of how it is specifically organized: (1) students receive quality instruction in general education settings; (2) student progress in the general curriculum is monitored; (3) students who do not make progress receive additional or alternative instruction; (4) these students' progress is again monitored; and (5) students who have still not demonstrated adequate progress are either evaluated for, or in some cases directly placed, into special education. These basic components of RTI have been echoed throughout more recent literature on the subject (e.g., Berkeley, Bender, Peaster, & Saunders, 2009; Marston, 2005; NCRTI, 2010). However, significant variation exists with regard to the procedures by which these steps are carried out in schools and the level of empirical support for these methods.

Quality instruction and progress monitoring in the general curriculum. RTI begins with the use of empirically-based instructional methods in general education classrooms and assessment of the degree of progress made by all students. Progress monitoring in general education settings can be accomplished in numerous ways. In determining whether or not a student is making progress in the general curriculum, school multidisciplinary teams may refer to data from homework assignments, classroom tests, school district assessments, and state standardized tests. Yet, a more popular and

arguably more psychometrically-sound method of assessing progress is through the use of curriculum-based measures (CBMs; Deno, 1985).

CBMs were originally conceived by Deno and Mirkin (1977) as a way for special education teachers to systematically trace student progress over time to inform data-based decision-making. CBMs are brief assessments of specific academic skills that are based directly on the material students are learning in the classroom. Although CBMs fit into the broad category of curriculum-based assessment, they are more proscribed in that they tend to consist of a few standardized tasks, have specific administration and scoring criteria, and are developed with concern for psychometric properties of reliability and validity (Deno, 2003). Although they began as a tool primarily used in special education settings, CBMs have since become popular for use with all students due to the ease of their use and the utility of the data they yield. CBMs typically take between 1 and 3 minutes for students to complete, and can feasibly be given by teachers, paraeducators, and even parents (Deno, 1985, 2003). They are specifically designed to be given multiple times across the school year with minimal modification, allowing for analysis of growth made by individual children and comparison across students (Tindal, 2013).

Extensive research has been done within the past three decades regarding how CBMs should be developed. According to Deno and Fuchs (1987), CBM paradigms for progress monitoring must include several core components: specification of the relevant outcome variable being measured, the stimulus material provided to the child, administration and scoring instructions, and information on how to use the raw data to make instructional decisions (i.e., norms and cutoff scores). CBMs are defined by their

direct connection to what students are learning in the classroom, which should allow teachers to inform their own instructional practices. With minimal training, teachers can create their own CBMs by pulling directly from classroom material (Shinn, 1989). However, they may also use published CBM protocols that have been developed with consideration of the typical curriculum at each grade level, such as the Dynamic Indicators of Early Literacy Skills program (DIBELS; Good & Kaminski, 2002) or the System to Enhance Education Performance model (STEEP; Witt, Daly, & Noell, 2000). Likewise, when using CBM data to make educational decisions about student progress, school personnel may refer to published national norms of student performance, or they may create their own local school or district norms once a large number of students have been assessed (Deno, 2003).

To date, CBMs have been developed to assess multiple skills in the areas of reading, writing, and mathematics – all core academic areas of learning disability identification. A brief summary of developments in each area is provided.

CBMs in reading. The body of research on reading CBMs is far more extensive than the research base for any other subject area. Measures of oral reading fluency, or the number of words a child is able to correctly read aloud in a given period of time (typically 1 minute), were among the first CBMs to be studied and remain the most popularly used for elementary school children (Tindal, 2013). However, additional measures of reading skills have since been developed that allow for assessment of early pre-reading skills, as well as advanced vocabulary and comprehension skills. For example, it is common for kindergarten and first grade students to be assessed using

measures of letter-sound fluency (the speed at which students can produce sounds associated with written letters) or word identification fluency (the speed at which students can read single words on a page) (Fuchs, Fuchs et al., 2012). Upper elementary and middle school students are commonly assessed using measures of vocabulary-matching (the speed at which they can match vocabulary terms to their definitions) or maze fluency (the speed at which they can, while silently reading a passage, select from a set of words or phrases which one would most appropriately fill in the blank in a sentence; Espin & Tindal, 1998; Fuchs, Fuchs et al., 2012). Performance on these measures reflects not only basic fluency, but higher level skills that become the focus of reading instruction in upper grade levels.

An extensive body of research supports the psychometric properties and utility of oral reading fluency CBMs. Early (Marston, 1989) and more recent syntheses of the literature (Reschly, Busch, Betts, Deno, & Long, 2009; Wayman, Wallace, Wiley, Tichá, & Espin, 2007) indicate that read-aloud passage probes show consistently high reliability ($r > .90$ in most studies) and can successfully predict a variety of variables ($r > .80$ in most studies), including individual and group reading achievement test scores. Scores on these measures have also been found to be predictive of classroom grades, special education placement, and teacher ratings of student performance (Fewster & MacMillan, 2002; Hintze, Shapiro, Conte, & Basile, 1997). Moreover, oral reading fluency has been found to predict reading comprehension performance as well as, if not better than, other measures of reading comprehension do (Fuchs, Fuchs, & Maxwell, 1988; Markell & Deno, 1997). Some cross-sectional work has indicated that oral reading fluency appears

to be a much better predictor of reading test scores for students in lower grades (i.e., grade 3 and below) than for students in upper grades (Hosp & Fuchs, 2005; Jenkins & Jewell, 1993). However, in their meta-analysis of 30 years of CBM studies, Reschly et al. (2009) found no statistically significant differences across grades1-6 in how well oral reading fluency and overall reading ability correlated, indicating that this issue is perhaps still unresolved.

Though relatively less research has been completed on maze CBMs, there is also evidence to support their reliability, validity, and overall utility. Given that maze CBMs can be administered to large groups of students at the same time via paper and pencil or computer, they may prove more time-efficient than other measures of reading comprehension (Fuchs & Fuchs, 1992). Although early research indicated that oral reading fluency performance was a significantly better predictor of overall reading ability than maze performance for students in early elementary grades, both had roughly equal predictive validity for fifth and sixth graders (Jenkins & Jewell, 1993). Subsequent comparative research has supported the psychometric superiority of oral reading fluency for assessing younger students, the comparability of the two types of CBM with upper elementary school students, and the superiority of mazes for tracking growth in middle school students (Ardoin et al., 2004; Graney, Martinez, Missall, & Aricak, 2010; Tichá, Espin, & Wayman, 2007).

The development and use of CBMs to assess pre-reading skills in young children for whom passages are inappropriate is an area of emerging research. Though evidence exists to support the reliability of basic word identification fluency, nonword fluency, and

letter-reading CBMs, word identification tasks have demonstrated significantly greater concurrent and predictive validity with first grade students (Daly, Wright, Kelly, & Martens, 1997; Fuchs et al., 2004). In a subsequent study, Compton, Fuchs, Fuchs, and Bryant (2006) used word identification, phonemic awareness, rapid naming, and oral vocabulary measures with children at the start of first grade, and found word identification to be the most useful metric for predicting at-risk status in reading at the end of second grade. Thus, for children who are not yet able to read full sentences, using simple word identification CBMs may be a viable option.

CBMs in writing. Although less research has been done on the use of writing CBMs, several types of writing measures have been developed and tested. The most basic writing measures involve requiring students to write for 3-5 minutes in response to a prompt (e.g., to complete a story that has been started for them) and then tallying the number of words written and/or the number of words spelled correctly (Amato & Watkins, 2011). More advanced scoring methods include counting the number of correct/incorrect letter sequences and correct/incorrect word sequences, which also take into account the child's use of appropriate spelling, grammar, capitalization, punctuation (McMaster & Espin, 2007). Alternatively, the percentage of these sequences out of the total number of sequences present in the student's writing sample may also be used (Tindal, 2013).

Research has indicated that, in general, the psychometric properties of writing CBMs vary depending on the grade level of the students one is working with, as well as scoring methodology. In a review of 28 studies of writing CBMs, McMaster and Espin

(2007) noted that while early studies of writing CBMs showed moderate to strong correlations between words written/spelled correctly and performance on other writing assessments (r = .67-.88), more recent studies have not been as positive, yielding significantly lower validity coefficients and providing insufficient evidence to support the reliability of these measures with elementary school children. The extant evidence for the use of writing CBMs with secondary students is better, although some researchers have expressed concerns that these measures are not sensitive enough to be used for routine progress monitoring (McMaster & Espin, 2007; Parker, Tindal, & Hasbrouck, 1991). With older students, it appears that using longer probes and more complex methods of scoring allow writing CBMs to better predict other writing markers. For example, recent work has supported the number of correctly placed punctuation marks, the number of correct or incorrect word sequences, and percentage of correct word sequences as the best markers of overall writing skill with secondary students (Amato & Watkins, 2011; Diercks-Gransee, Weissenburger, Johnson, & Christensen, 2009; Lopez & Thompson, 2011). Current studies (Campbell, Espin, & McMaster, 2013; Espin et al., 2008) have also demonstrated greater reliability and validity for probes of at least 7 minutes over the traditional 3-5 minute measures.

CBMs in mathematics. CBMs have also been developed and studied to assess skills in mathematics. Overall, math CBMs are arguably more diverse in their content than reading and writing measures because of the diversity inherent in the core math curriculum across grade levels. Thus, math CBMs for very young students tend to target early numeracy skills and counting, probes for elementary students typically involve

number identification and basic computation, and measures for older students may incorporate advanced concepts such as algebra (Lembke, Hampton, & Beyers, 2012). In the majority of math CBMs, students are given 2-8 minutes to write the answers to as many math problems as possible (Christ, Scullin, Tolbize, & Jiban, 2008). They are usually scored by counting the number of correct digits in the correct places that a student provides when responding to the problems (Hosp & Hosp, 2003); consequently, although the student's overall answer to a two-digit problem may be incorrect, he or she can still receive some credit if one digit was correct. In recent years, researchers have explored computer-based administration of math CBMs. Though this method shows promise, caution is recommended because results may not generalize from paper-and-pencil to computer modalities (Duhon, House, & Stinnett, 2012).

Evidence for the use of math CBMs is similar in scope to the evidence for writing CBMs. Test-retest reliability evidence for math CBMs was lacking in the literature until recently, but inter-rater and alternate form reliability estimates have consistently been above .80 (Christ et al., 2008; Foegen, Jiban, & Deno, 2007). Validity evidence has been more limited. Marston's (1989) early review of the literature on this topic yielded validity coefficients with other measures in the .26-.67 range, and although more recent reviews by Foegen et al. (2007) and Christ et al. (2008) have pointed to more positive results, these measures still tend to demonstrate less predictive capability than reading CBMs. Although Christ et al. (2008) found median correlation coefficients between CBMs and other measures of math ability in the .74-.83 range, the authors cautioned that these results were likely inflated due to the statistical issues inherent in using cross-grade

samples, and argued that math CBMs are prone to construct underrepresentation because they assess such a narrow range of skills. More recent studies have supported the theme that math CBMs tend to show strong reliability and moderate levels of both concurrent and predictive validity, but that overall, they remain less psychometrically sound than reading CBMs (Clarke et al., 2011; Seethaler & Fuchs, 2010, 2011).

Incorporating progress monitoring into decision-making. As students' progress in the general curriculum is monitored through the use of curriculum-based measures, extensive amounts of data are generated for school teams to analyze and incorporate into decision-making. Yet disparities exist, both within the RTI research community and the schools, with regard to how frequently measures should be administered, how many measures should be taken into account, and what level of progress should imply that a student requires intervention.

Universal screenings. A core tenet of RTI paradigms is that they reach all students in a school, not merely those with readily identifiable learning difficulties. Consequently, universal screenings of every student in a school in core academic areas using CBMs at grade level are the foundation of RTI (Fuchs & Fuchs, 2005; Fuchs, Fuchs et al., 2012). Screenings occur during the fall (typically the first month of the school year) and spring, and in some cases, the winter as well (Fuchs & Fuchs, 2006; Jenkins, Hudson, & Johnson, 2007). It is recommended that the same measure be used across screenings in a single school year to prevent any confounding effects of potential variations in probe difficulty (Ardoin & Christ, 2008). Although some have suggested that the practice of administering three measures to each student and taking the median

value is ideal (e.g., Shinn, 2002), research has indicated that use of a single probe is sufficient for screening purposes and does not significantly reduce the amount of variance explained in a student's achievement (Ardoin et al., 2004).

Universal screenings are thought to prevent students who are struggling from going unnoticed. Research indicates that, while teachers can effectively identify students who would benefit from intervention through a basic referral process, the simple referral process often misses students with significant difficulties (VanDerHeyden & Witt, 2005). Even so, universal screenings are not without criticism. Recently, Fuchs, Fuchs et al. (2012) have argued that one-stage universal screenings may lead to excessive "false positives", or students who appear to be at risk, but are truly not. Thus, Fuchs and colleagues recommend using a two-stage screening process, with the first stage involving a reading screener with a very high cutoff point passable only by students who are clearly not at risk, and the second stage including additional assessment to distinguish false positives from students who truly require help.

Current research is providing initial support for the notion of two-stage universal screenings. Compton et al. (2010) used this model, with stage 2 constituting 5 weeks of progress monitoring in word identification fluency for one group of students, and one 20-30 minute "dynamic assessment" for a second group. Dynamic assessment (Feuerstein, 1979; Grigorenko & Sternberg, 1998) involves quantifying the amount of scaffolding or prompting that is required for the child to complete a particular task. Both use of the 5-week progress monitoring and the dynamic assessment were able to significantly reduce false positives over a one-stage model (Compton et al., 2010). A follow-up study using a

similar process with math similarly demonstrated the superiority of a two-stage model with dynamic assessment over a single universal screening (Fuchs et al., 2011). Most recently, Fuchs, Compton et al. (2012) used a two-stage screening process with first graders, with the second stage consisting of measures of four cognitive dimensions (rapid naming, phonological processing, listening comprehension, and nonverbal reasoning). After tracking students through fifth grade to see who were later diagnosed with reading disabilities, they concluded that the two-stage screener reduced the number of false positives, or students who would have been unnecessarily placed in intervention, significantly over a one-stage process.

Using data to identify struggling students. Different recommendations also exist with regard to how students should be identified as "struggling" based on universal screening data, and what percentage of students in a school should be considered needing intervention. Adopting a systems perspective, Ikeda, Neessen, and Witt (2007) have argued that a strong general education curriculum should meet the needs of 80% of learners, meaning that about 20% of students will fall below grade level and require extra support. Fuchs and Fuchs (2005) recommend that students who fall below the 25th percentile in their school based on universal screening data be considered at risk of school failure. However, Fuchs and Fuchs (2006) also argue that best practice is to monitor the progress of these students in the general curriculum for several weeks after the screening using CBMs, and to compare the progress they make to normative or criterion-referenced standards for improvement to determine if they are making progress in the general education curriculum before considering intervention.

A supplement or alternative to the approach advocated by Fuchs and Fuchs (2006) is the use of a skill versus performance deficit assessment post-screening. Also known as "can't do/won't do" assessments, these measures can facilitate intervention selection by helping to determine whether a student's poor performance in a screening is due to a lack of motivation or a true skill deficit (VanDerHeyden & Witt, 2008; Witt & Beck, 1999). Students are administered the same CBM again, but offered reinforcement contingent on improving their previous performance. If measurable improvement is demonstrated, it may be assumed that the students' low screening results are due, at least in part, to motivation. There is no set standard for defining marked improvement, but growth of 20-50% over the previous score and, in some cases, a second score within the instructional range have been used in the literature (Ardoin, Witt, Connell, & Koenig, 2005; Duhon et al., 2004; Noell, Freeland, & Witt, 2001).

Relatively little research has been completed on the incorporation of skill/performance assessments into RTI models. Noell et al. (1998) demonstrated that providing reinforcement alone could increase oral reading fluency scores in some students. Building upon these initial findings, Duhon et al. (2004) screened students in reading and completed a skill/performance deficit assessment with low scorers. They then implemented both instruction-focused and motivation-focused interventions, and found that students made the most growth when the intervention was matched to the results of the skill/performance assessment (i.e., a motivational intervention for a student with a student hypothesized to have a motivational issue; Duhon et al., 2004). Additional studies have indicated that, if about 15% of students meet criteria for being at-risk based

on screening data, between one-third and one-half of those students may display performance deficits when administered can't do/won't do measures (VanDerHeyden et al., 2003; VanDerHeyden, Witt, & Gilbertson, 2007). Thus, incorporating this step into the universal screening process has the potential to refine the RTI process by ensuring that students are directed to the intervention – academic or motivational – that is most appropriate for their needs.

Implementing interventions and monitoring progress. After students have been identified as at-risk of failure in one or more academic domains, they are ideally placed into interventions deemed appropriate for their needs. Regardless of whether a standard protocol or problem solving approach is taken, it is essential that interventions be evidence-based and delivered with integrity and fidelity, and that progress monitoring data be continually collected (Fuchs et al., 2003; Fuchs, Fuchs et al., 2012). However, there is variation within the RTI field with regard to how interventions are implemented and how progress monitoring data is used to inform further decision-making.

Making decisions regarding student progress. Significant disparities also exist in the methods that practitioners use to determine whether or not a student is making progress in intervention. A student's scores across time on regularly administered CBMs are typically plotted in graphical form, along with a goal line to demarcate the ideal rate of progress based on normative data (Ardoin, Christ, Morena, Cormier, & Klingbeil, 2013). Often, school teams use professional judgment to consider these graphs and make decisions about the intensity, frequency, and content of interventions. However, Ardoin et al. (2013) note that one of two decision rules may be used to determine whether or not

adequate progress has been made – the data point rule or the trend line rule. Practitioners using a data point rule consider the child as having made progress if the last few data points (typically 3-5) hover about the goal line; if points are continually far below or above the line, the intervention is adjusted to better meet the student's needs (Ardoin et al., 2013). Alternatively, practitioners employing a trend line rule use statistical procedures to transform the student's data points into a trend line (Christ, Zopluoglu, Long, & Monaghen, 2012; Parker & Tindal, 1992). The focus with this method is the slope of that line, which is compared to the slope of the goal line to inform intervention decisions (Deno, Fuchs, Marston, & Shin, 2001). In lieu of normative data to construct a goal line, practitioners may use the standard of a trend line slope greater than the standard error of estimate to determine the presence of progress (Fuchs & Fuchs, 2005).

Debate exists with regard to the length of intervention and amount of data necessary for measurable, reliable improvement to be seen. It is recommended that progress monitoring take place at least monthly for students receiving tier 2 supports, although often students are monitored more frequently (Burns & Coolong-Chaffin, 2006). In reviewing the past 50 years of literature on this subject, Ardoin et al. (2013) found that most works referenced used 6-10 data points for decision-making. However, research has demonstrated that using such a limited number of data points often leads to error in prediction with trend lines. For example, when Shinn, Good, & Stein (1989) used 10 data points (biweekly progress monitoring over 5 weeks) to predict oral reading fluency 6 weeks later, predictions were off by an average of 18 words per minute – a significant figure. Good and Shinn (1990) later found that using 20 data points instead of 10

significantly reduced error rates. Yet, more recent research has indicated that, even with that many data points, there may still be a significant amount of error in prediction because student growth is such an unstable construct, and because progress may also be a function of the specific passage material used (Ardoin & Christ, 2009; Christ, 2006).

All in all, the existing research on decision rules has yet to provide extensive support for one decision method over the other. Ardoin et al. (2013) reviewed more than 100 works on decision rules pertaining to oral reading fluency CBMs, and determined that, while only a few referenced the data point rule exclusively, nearly half referenced only the trend line method or both methods. Notably, no published study provided evidence for the decision accuracy of any decision rule alone, likely because decisions about responsiveness are usually made in conjunction with other information in addition to progress monitoring data (Ardoin et al., 2013).

Defining typical growth. Despite the present ambiguity regarding the accuracy of decision rules in progress monitoring, the literature on this subject has yielded important normative data on student growth rates. Although such data should not negate the necessity of developing local norms (Deno, 2003) and focusing on children as individuals (Fuchs, Fuchs et al., 2012), it can facilitate decision-making for practitioners in addition to or in the absence of local data.

Just as CBMs for reading fluency have received the greatest focus in the literature, so has the collection of normative data in reading. In analyzing large, nationally representative datasets of oral reading fluency CBM scores, both Fuchs, Fuchs, Hamlett, Waltz, and Germann (1993) and Deno et al. (2001) concluded that students

making adequate progress can reasonably be expected to gain about 2 words per week in first grade, around 1.5 words in second grade, and approximately 1 word per week in grades 3-6. Rates for students with identified learning disabilities are markedly lower, at .8 words per week in first grade and .5 words per week in grades 2-6 (Deno et al., 2001). Subsequent studies have supported the general downward trend in oral reading fluency growth across the elementary and middle school years, and have indicated that growth tends to be greater in the fall semester than in the spring semester, particularly for general education students (Ardoin & Christ, 2008; Christ, Silberglitt, Yeo, & Cormier, 2010; Hasbrouck & Tindal, 2006). However, other research has suggested that growth is actually greater in the latter half of the school year (Graney, Missall, Martinez, & Bergstrom, 2009).

Relatively less research has been completed to identify and confirm typical growth rates in mathematics and writing. Unlike reading growth rates, math CBM growth rates tend to increase through the late elementary school years, with students improving by .3 digits correct per week in grades 1-3, .70-.75 digits per week in grades 4-5, and .45 digits per week in grade 6 (Fuchs et al., 1993). Conversely, writing CBM scores display a pattern similar to that of oral reading fluency, with growth in the number of correct letter sequences declining geometrically across the elementary years, from about 1 letter per week in grade 2 to about .2 letter per week in grade 6 (Fuchs et al., 1993). More recent work has shown that math growth tends to be greater from winter to spring than from fall to winter (Graney et al., 2009). Follow-up work to confirm trends across grade levels in writing CBMs has yet to be published.

Meeting the needs of nonresponders. Currently, disagreement exists within both the practitioner and research communities with regard to how to best meet the needs of nonresponders within an RTI framework. Two major areas of debate have been the role that tier 2 interventions should play for students displaying the most severe of learning problems, and the design of tier 3 levels of support.

The necessity of tier 2 intervention. One emerging issue in the RTI literature is whether or not tier 2 intervention is necessary for all struggling students. That is, if a student is struggling in the general curriculum to an extreme degree, is a lower level of intervention necessary, or should the student begin with the most intensive form of intervention available (which, in some models, constitutes special education evaluation and placement)? As Fuchs, Fuchs et al. (2012) point out, correctly identifying students who require the highest level of service both prevents the students from struggling for an extended period of time and helps conserve financial resources in a school. It may also help districts adhere to IDEIA's (2004) Child Find provisions, which require schools to locate, identify, and evaluate children with disabilities in a timely manner.

Unfortunately, as of yet, no clear method exists for successfully identifying nonresponders before they have been given a chance to respond. However, Compton, Gilbert et al. (2012) successfully used logistic regression to predict whether or not students would respond to tier 2 intervention. Using universal screening data, progress monitoring data from the general curriculum (tier 1), teacher ratings of attention and behavior, and performance on selected subtests from standardized assessments, Compton and colleagues correctly identified 90% of the students who would be nonresponsive to

tier 2 intervention. Perhaps, even more important is the fact that relying solely on screening data did not significantly alter this figure, although it did reduce specificity (the ability to correctly identify students who would respond; Compton, Gilbert et al., 2012). Thus, there is initial evidence to suggest that "fast-tracking" students to more intensive intervention may be possible.

Procedures for tier 3 service delivery. A second major issue is determining what should constitute tier 3 services. A recurring theme in the RTI literature is that approximately 5% of all students (excluding those with known intellectual disabilities) will not respond to targeted tier 2 intervention, even if it is delivered consistently and with fidelity (Fuchs, Fuchs et al., 2012). Empirical studies of RTI systems have largely supported this estimate (see Burns, Vanderwood, & Ruby, 2005; Fuchs, Compton, Fuchs, Bryant, & Davis, 2008; Vaughn et al., 2010). Yet, there is great variability in terms of how tier 3 services are implemented and conceptualized. Two major alternatives have emerged for tier 2 nonresponders – providing a more intensive level of intervention beyond tier 2 without formal special education, or evaluating and considering students for special education services.

Tier 3 services have been described as intensive and highly individualized interventions tackled from a problem-solving perspective. Though tier 2 interventions are typically delivered in general education settings, tier 3 interventions may be delivered by general or special education personnel to smaller groups of just one or a few students, and typically involve even more frequent progress monitoring (Burns & Coolong-Chaffin, 2006; Gersten et al., 2009). Burns and Coolong-Chaffin argue that the

distinguishing feature between special education and tier 3 intervention is the level of resources needed for the student to make ongoing progress – if the service becomes extensive, special education resources may need to be formally applied to educate the student successfully. To this point, research indicates that schools with both tier 2 and tier 3 intervention services often accomplish the increased intensity of tier 3, not just by increasing the time students spend in intervention, but by providing more positive corrective feedback, allowing students more response opportunities, slowing instructional pace, using more focused goals, and reducing the number of transitions between activities (Mellard, McKnight, & Jordan, 2010).

Relatively little research has been completed on the logistics and efficacy of providing tier 3, non-special education services to tier 2 nonresponders within an RTI framework. In most cases, however, tier 3 intervention has successfully remediated difficulties in at least some youth. Studies of tiered reading supports for early elementary aged children by O'Connor (2000), Berninger et al. (2000), and Vaughn et al.(2003) have all indicated that children who are extremely slow to respond to initial intervention efforts can make progress, and that for some, the improvement is enough to allow them to thrive in general education settings. For example, in the Vaughn et al. (2003) study of children at-risk of reading failure, 53% of students met exit criteria (grade-level reading proficiency) from tier 2 intervention, but another 22% were able to meet exit criteria after additional, more individualized tier 3 intervention. More recent work with middle school students has led to similar results, although the proportion of students able to attain grade-level reading proficiency after tier 3 intervention has been significantly

lower (Vaughn et al., 2010, 2012). Even so, these findings provide some empirical basis for conceptualizing tier 3 as an extension of tier 2.

Alternative models of RTI also consider tier 3 as a form of more intensive service delivery, but include it under the umbrella of special education services. In such models, when children fail to respond to tier 2 interventions, they are evaluated and considered for special education services (Fuchs & Fuchs, 2005; Fuchs, Fuchs, & Stecker, 2010). Yet even when tier 3 is equated with evaluation for special education services, disparities exist with regard to how this process can and should proceed.

With the advent of IDEIA 2004, schools are legally permitted to use a student's failure to respond to empirically-based intervention as the sole determinant in whether or not that student has a learning disability. As of 2010, 13 states required use of RTI in SLD identification (Zirkel & Thomas, 2010a), and six explicitly banned use of a discrepancy model (Zirkel & Thomas, 2010b); since then, Wisconsin has joined the list, bringing the total to 14 (Wisconsin Department of Education, 2011). Many more states encourage use of RTI data in identification and are in the process of developing their own RTI models (Berkeley et al., 2009; Zirkel & Thomas, 2010a). In some school systems, such as the well-known Heartland educational system in Iowa and the Minneapolis Public Schools, RTI-only practices have become well-established, and students are placed in special education after persistent failure to respond to intervention, but without a formal psychoeducational evaluation (Fuchs et al., 2003).The RTI-only approach is based on the fundamental assumption that, if students are unable to make progress after extensive adaptations to the general curriculum, then they must have an inherent deficit that can be

conceptualized as a learning disability (Vaughn & Fuchs, 2003). In schools with this approach, students are carefully progress monitored as they receive increasingly intense interventions, and a "dual discrepancy" model is often used that does not involve intellectual assessment – rather, those students who are both significantly below peers and making extremely poor progress may be identified as having learning disabilities (Shinn, 2007).

Yet many in the field have argued that before any student – even one who has been deemed nonresponsive to intervention – should be considered eligible for and in need of special education services, a multifaceted psychoeducational evaluation must take place involving cognitive and academic assessment. As Hale et al. (2006) note, nonresponsiveness alone does not address the legal definition of a learning disability as a "disorder in one or more… basic psychological processes" (IDEIA, 2004, 20 USC § 1401(30)). There can be more than one reason for a student's failure to respond, and the reason is not always SLD (Kavale et al., 2005; Reynolds & Shaywitz, 2009). Moreover, because IDEIA (2004) requires that students be assessed in all areas related to a suspected disability, Hale, Naglieri, Kaufman, and Kavale (2004) argue that it is necessary to investigate a student's cognitive strengths and weaknesses if SLD is suspected. For these scholars, completing a full psychoeducational evaluation is believed to be the only way to both adhere to legal requirements and to truly differentiate between students with learning disabilities (i.e., students with truly unexpected learning failure given their cognitive ability) and those who may simply be low achieving for other reasons.

A research base has emerged connecting performance in various cognitive subdomains to academic underachievement in specific areas (e.g., Compton, Fuchs, Fuchs, Lambert, & Hamlett, 2012; Evans, Floyd, McGrew, & Leforgee, 2002; Floyd, Evans, & McGrew, 2003; Floyd, McGrew, & Evans, 2008; Johnson, Humphrey, Mellard, Woods, & Swanson, 2010). Scholars have developed multiple methods of conducting comprehensive intellectual assessments that take into account cognitive strengths and weaknesses in SLD assessment. These include the discrepancy/consistency model (Naglieri & Das, 1997), concordance-discordance model (Hale et al., 2008), and the most well-known, cross-battery assessment (Flanagan et al., 2013). All have in common a focus on identifying areas of deficit in students' cognitive profiles that may explain their academic underachievement and, in theory, inform instruction. However, these approaches have been questioned based on concerns that they will lead to overidentification of children as learning disabled (Stuebing, Fletcher, Branum-Martin, & Francis, 2012). They have also been criticized on the basis of research suggesting that general intellectual ability, rather than performance in specific cognitive domains, is the best predictor of performance in academic areas (Glutting, Watkins, Konold, & McDermott, 2006; Parkin & Beaujean, 2012).

Overall, the issue of RTI models and comprehensive psychoeducational evaluations presently remains unresolved. In 2010, 58 major scholars in the learning disabilities field came together to produce a white paper consensus in conjunction with the Learning Disabilities Association of America (LDA; Hale et al., 2010). Drawing together major findings on this topic from the past few decades, the scholars produced a

set of recommendations that included maintaining the present conceptualization of SLD as a disorder of psychological processes and asserting that neither a simple discrepancy model nor RTI alone were enough to effectively identify students with learning disabilities. They advocated for the use of empirically-supported RTI models and comprehensive assessment of cognitive strengths and weaknesses for students who fail to respond. Importantly, Hale and colleagues stressed the need for greater research into precisely how knowledge of a child's strengths and weaknesses can inform instructional methods and enhance student achievement.

Critiquing RTI and the future of RTI models. The focus on RTI within the literature and in practice grew substantially after its explicit inclusion in IDEIA as a method by which children may be identified as having learning disabilities. However, RTI has been considered by many scholars to be revolutionizing education, not just for students with special needs, but for all children.

RTI models have not been without criticism. Although some early concerns about RTI have since been largely disconfirmed (e.g., the worry that adding the RTI provision to IDEIA (2004) would lead to a greater number of students being identified as having learning disabilities), others have remained (U.S. Department of Education, 2006). Some concerns have been addressed in litigation at the district and state levels, most of which have involved assertions that RTI models can delay the special education evaluation process (Yell & Walker, 2010; Zirkel, 2013). In general, the courts have ruled in favor of school districts, provided that they implemented RTI models while still abiding by the procedures outlined in IDEIA (e.g., a parent's right to request an

evaluation at any time; Yell & Walker, 2010). Additionally, some have argued that, due to the government provision for allocating some special education funding to RTI, these systems are often perceived as the responsibility of special education departments, when in principle they can and should be integral to general education practice (Wixson, 2011). Related to this is the overall lack of consistency in RTI models and how they are implemented across states, school districts, and even individual schools within a single district. Studies of this variability have revealed that RTI systems vary extensively in the number of tiers used, the specific CBMs and interventions implemented, and the way in which they incorporate RTI data into psychological evaluations (Berkeley et al., 2009; Jenkins, Schiller, Blackorby, Thayer, & Tilly, 2013). Scholars have expressed concern that overreliance on RTI data may prevent school psychologists from performing in-depth cognitive and neuropsychological assessment that has the potential to inform remediation of learning difficulties (Hale et al., 2010; Kavale et al., 2005).

Despite these criticisms, RTI has received much praise for its congruence with the modern educational reform movement. Both NCLB (2001) and IDEIA (2004) stressed the importance of approaching education from a prevention and early intervention framework, and RTI contributes to this goal in several ways. The U.S. Department of Education (2006) has specifically asserted that RTI data can help determine whether a child's low achievement could be due to a lack of appropriate instruction – a rule-out that would prevent the student from being considered learning disabled under the law, and a key factor in helping schools improve the quality of instructional delivery for every student. Researchers have also suggested that RTI models allow for higher degrees of

ecological validity than traditional assessment methods, meaning they provide a better connection between the skills being assessed by the school psychologist and the skills students need to succeed in the daily classroom setting (Dean, Burns, Grialou, & Varro, 2006). Moreover, as Fuchs et al. (2003) argue, RTI is thought to save school districts money because, by identifying students who are struggling early on and intervening before they fall too far behind, it reduces the number of children who are falsely identified as having disabilities. This should lead to lower numbers of students enrolled in special education services, and studies indicate that it does (e.g., O'Connor, 2000; VanDerHeyden et al., 2007). Furthermore, in schools with RTI frameworks, students identified as learning disabled tend to be more impaired than those identified in schools without RTI frameworks, suggesting that RTI promotes identification of students for special education whose challenges are the most difficult to remediate (O'Connor, Bocian, Beach, Sanchez, & Flynn, 2013).

Although RTI remains a contested topic within the field, it holds promise for changing the way we think about education and making our educational system more equitable and effective for every child. A major issue that has emerged in the decades since the passing of the original EAHCA is the disproportional representation of students from minority backgrounds in special education (Donovan & Cross, 2002; Shifrer, Muller, & Callahan, 2011).The issue is particularly pertinent for students who are learning English as their second language, because they typically face both ethnic and linguistic minority status within the American educational system. Some have argued that RTI models of service delivery have the potential to reduce this disproportionality

and create more equitable outcomes for all students, regardless of their ethnic or linguistic background.

Educating CLD Students

As education has moved toward developing and refining RTI in recent years, the composition of the U.S. public education system has also been evolving. CLD students have come to represent one of the fastest-growing segments of the K-12 student population in this country (Fry, 2007). The term CLD has emerged in recent years as an alternative to the older "English language learner" (ELL) or "limited English proficient" (LEP) terms, and is now often used interchangeably with these terms in the literature (e.g., Hart, 2009; Lakin, 2012). Watson, Rinaldi, Navarrette, Bianco, & Samson (2007) of the Council for Learning Disabilities provide the following definition of a CLD student:

> The term 'cultural and linguistic diversity' stresses the relationship between language and culture and the characteristics of students who are ethnically, racially, culturally, and linguistically different from the mainstream population. A culturally and linguistically diverse (CLD) student, or English language learner (ELL), is one who has to acquire a second or additional language and culture, a process that can be very challenging ("Who is the culturally and linguistically diverse learner?", para. 1).

In practice, many states still use the term ELL when referencing the type of services delivered to children who are identified as not being English proficient (Goldenberg & Rutherford-Quach, 2012; Ragan & Lesaux, 2006). Although students

with limited English proficiency have been part of the nation's school system since its inception, recent demographic shifts have greatly enhanced the need for ways to effectively educate these students. NCLB (2001) formally called for schools to begin closing the notable achievement gap between CLD and non-CLD students, and stipulated that state achievement test data must be reported for this subgroup. Within the field of school psychology, similar concerns have been expressed regarding the need for a greater focus on training school psychologists in how to work with CLD populations (Lopez & Bursztyn, 2013).

CLD students in U.S. schools. According to the U.S. Department of Education's National Center for Education Statistics (NCES; 2012), CLD students comprise roughly 10% of the total population of K-12 students in the United States – an increase of 2% since 2000. The proportion of students in the United States who are considered CLD is expected to grow significantly in the near future, reaching an estimated 40% by the year 2050 (Goldenberg, 2008). Additionally, though not all are formally classified as CLD students, one in five K-12 students in the United States comes from an immigrant family, and 78% of these children speak a language other than English at home (NCES, 2012). Among families who do not speak English in the home, Spanish is by far the most commonly spoken language, followed by Chinese, French, and Tagalog (U.S. Census Bureau, 2010). The distribution of CLD students in schools is uneven across the nation, with the greatest concentrations in urban areas and the Southwest (NCES, 2012). The schools with the highest numbers of CLD students are also more likely to have higher

student-to-faculty, more students qualifying for free and reduced-price lunch, and Title I status (Fry, 2008).

Research indicates that CLD students face significant challenges within the American educational system. CLD students tend to score significantly lower than monolingual students on standardized tests of academic skills, particularly in reading. For example, data from the National Assessment of Educational Progress (NAEP) indicated that, among CLD fourth graders, 73% failed to achieve proficiency in reading and 46% failed to achieve proficiency in math, compared with 25% and 11% of non-CLD students respectively (Fry, 2007). The achievement gap between CLD students and their peers becomes even greater in middle school (Fry, 2008). CLD students are also much more likely to be retained in school for academic difficulties than are English-proficient students (Wang & Wang, 2007), which is especially problematic given the wealth of research on the negative consequences of grade retention (see Jimerson, 2001 for a meta-analysis). As secondary students, CLD students are three times as likely to drop out of school as non-CLD students, and the risk is far greater for those who have the weakest English language skills (August & Shanahan, 2006).

Just as the proportion of CLD students varies widely from state to state, so do the methods by which states address the needs of these students. States are not specifically required to provide bilingual education, but the Civil Rights Act (1964) prohibits schools from denying access to education based on race, ethnicity, or national origin. In 1974, the Supreme Court ruling in *Lau v. Nichols* set a precedent with the idea that linguistic proficiency should be treated as a proxy for national origin, and thus schools must take

steps to remove barriers to accessing the curriculum that stem from limited English proficiency (Jacob & Hartshorne, 2007). States are currently provided funding to implement instructional strategies that promote English proficiency through NCLB (2001), but how they must do so is not specified in the law. Multiple models of ELL instruction exist, including structured English immersion (in which students are placed in self-contained classrooms and taught academic subjects solely in English), English as a second language (in which students receive pull-out support in their native language, and bilingual education (in which students are taught in both their native language and English; Honigsfeld, 2009). Research indicates that bilingual instruction fosters academic achievement in children's second language (August & Shanahan, 2006; Rolstad, Mahoney, & Glass, 2005; Slavin & Cheung, 2005). However, all-English instruction has become the most common method of instruction for this population, even as support for the efficacy of this approach is extremely limited (Goldenberg, 2007; Zehler et al., 2003).

Determining linguistic proficiency. The methods by which students are identified by schools as in need of supplemental English instruction are similarly variable across states and have also been called into question. In most states, schools administer a home language survey to parents when children are first enrolled, and the results of this survey determine the need for further assessment and ELL instructional placement (Goldenberg & Rutherford-Quach, 2012; Ragan & Lesaux, 2006). Although some surveys simply ask what language is primarily spoken in the home, others are more complex and include questions about the child's first language or the languages that are

used by specific members of the household (Bailey & Kelly, 2010; National Research Council [NRC], 2011). Unfortunately, the research on whether or not these surveys effectively identify students in need of ELL instruction is extremely limited, and suggests that use of single-item home language surveys leads to underidentification of CLD students for ELL services (Bailey & Kelly, 2012; Goldenberg & Rutherford-Quach, 2012).

Once students are identified as potentially in need of ELL instruction, further assessment is done to determine their initial proficiency levels and track their progress in language acquisition over time. NCLB (2001) requires each state to establish its own standards for levels of English proficiency, and these standards inform the development of proficiency tests. Many measures are thus state-specific, but may be based on widely available, nationally-normed tests such as the Stanford English Language Proficiency Test (SELP; Harcourt Assessment, 2004), Language Assessment Systems Links (LAS Links; DeAvila & Duncan, 1990), Woodcock-Muñoz Language Survey Normative Update - Revised (WMLS-R NU; Woodcock & Muñoz-Sandoval, 2011), and Bilingual Verbal Ability Tests – Normative Update (BVAT -NU; Muñoz-Sandoval, Cummins, Alvarado, & Ruef, 2005). The reliability and validity of the state-specific, modified forms of published measures, including the California English Language Development Test (CELDT; California Department of Education, 2008) and Arizona English Language Learner Assessment (AZELLA; Arizona Department of Education [ADE] & Harcourt Assessment, 2007), has been heavily questioned (Bailey & Huang, 2011; NRC, 2011). Studies comparing student performance on multiple proficiency tests have indicated that

the results are often inconsistent across measures, and that the tests frequently underestimate the true proficiency of students (MacSwan & Rolstad, 2006; Pray, 2005).

Additional criticism has been leveled at the criteria schools use in determining whether or not a student should be exited from ELL services and reclassified as English-proficient. A review of practices across states by Wolf et al. (2008) indicated that while some states use just a single proficiency test score, others use as many as six different criteria, including multiple tests, student portfolios, and input from parents and teachers. As with the research on home language surveys, studies of ELL exit criteria imply that more information leads to better decision-making. In investigating Arizona's change from relying on multiple measures for reclassification to a single measure (the SELP), Mahoney, Haladyna, and MacSwan (2009) found that use of just one tool led to a greater number of students being exited from ELL instruction than did the use of multiple measures, and speculated that, for many of these students, the exit was premature. This finding was supported by Garcia, Lawton, and Diniz de Figueiredo (2010), who also studied Arizona's proficiency exam and found that it tended to overpredict academic achievement, particularly for middle school students. Taken together, these findings imply the existence of a significant gap between research and practice with regard to proficiency testing practices.

CLD students and special education. The issue of how to best educate CLD students with special needs is even more complex. IDEIA (2004) specifies that a lack of English proficiency must be ruled out as the sole reason for a student's academic difficulties when considering special education services, but it is far less specific on how

schools should meet the needs of CLD students in special education. As Klingner, Boelé, Linan-Thompson, and Rodriguez (2014) note, CLD students with special needs benefit from teachers who are culturally and linguistically responsive, instructional strategies that are empirically-validated for use with CLD students, and ongoing support to help them acquire English. Even so, a majority of states do not address CLD students specifically in their special education laws, beyond the information inherent in federal legislation (Scott, Boynton Hauerwas, & Brown, 2014). Similarly, approximately 75% of school districts across the United States do not have programs specifically designed for these students; rather, staff members from CLD and special education programs provide services in collaboration with one another (Zehler et al., 2003).

More than half of all CLD students who are referred for special education services are referred for reading difficulties (U.S. Department of Education, 2003). Research shows that while CLD students tend to be generally overrepresented within the SLD, speech-language impairment, and mild intellectual disability categories, they are underrepresented within the emotional disturbance category (Sullivan, 2011). Within the SLD classification, CLD students who have limited proficiency in both languages they speak constitute an even more disproportionate subgroup (Artiles, Rueda, Salazar, & Higareda, 2005). However, the issue is more complex than it initially appears. Linguistic disproportionality tends to be more pronounced in wealthier schools than poorer ones (Argulewicz, 1983). Additionally, some research implies that CLD students may actually be underrepresented in special education in the earliest grade levels, and that overrepresentation does not emerge until third grade (Samson & Lesaux, 2009).

Traditional assessment of SLD in CLD students. To date, the research on specific assessment procedures utilized with CLD students suspected of having SLD is somewhat limited. In the late 1990s, multiple publications emerged from a national survey-based research project designed to study the practices of school psychologists with CLD and bilingual students. Ochoa, Powell, and Robles-Piña (1996) examined the specific assessments used by school psychologists with the bilingual and limited English proficient (LEP) population and, in relating their results to a comparable study of practices with the general population of students (Stinnett, Havey, & Oehler-Stinnett, 1994), concluded that psychologists tended to use a more varied set of assessments with bilingual and LEP students than with the native English-speaking population. The most frequently utilized cognitive assessments were the Wechsler Intelligence Scales for Children, 3rd Edition (WISC-III), Draw-a-Person Test, and Leiter International Performance Scale, all given in English; the latter two and the WISC-III Performance Scale were considered to be nonverbal measures, although instructions may have been given in English (Ochoa, Powell et al., 1996). Academic achievement testing was more commonly given in English than Spanish, although a significant number of practitioners reported use of Spanish achievement tests, including the Batería Woodcock Psico-Educativa en Español and the Brigance Diagnostic Assessment of Basic Skills – Spanish Version (Ochoa, Powell et al., 1996). A majority of school psychologists who conducted bilingual assessments reported having used interpreters during the process, but only a fraction of these interpreters had ever been formally trained (Ochoa, González, Galarza, & Guillemard, 1996). These results were highly consistent with earlier, smaller-scale

studies on the same topic (see Morris, 1977; Nuttall, 1987), indicating that very little had changed in terms of common practices with this population over the decades.

Researchers have also explored the specific methods and rationale by which school psychologists rule out linguistic proficiency as a factor in disability determination. As part of the aforementioned survey-based project, Ochoa, Rivera, and Powell (1997) asked psychologists to report how they addressed this exclusionary stipulation, and found that very few considered the number of years of instruction in English that a student had received, or whether or not an IQ/achievement discrepancy existed in both the child's native language and English. Rather, the most commonly used pieces of information were parent input on the student's language use, the length of time the student had lived in the United States, educational history, and state language proficiency test scores (Ochoa et al., 1997). More recent research on psychoeducational evaluations of CLD students with SLD has demonstrated that, in many cases, the students do not actually meet criteria for SLD because the impact of linguistic proficiency was never properly ruled out (Wilkinson, Ortiz, Robertson, & Kushner, 2006). In the Wilkinson et al. (2006) study, for example, an in-depth review of cases of CLD students who were classified as SLD revealed that roughly half of these students had not been in school consistently and had not been exposed to early intervention – factors that undoubtedly affected their academic achievement and may have made them falsely appear to be learning disabled.

All of these findings prompt questions regarding the discrepancy between actual CLD evaluation practices in the schools, and best practice according to legal and professional standards. Federal law requires the following:

> Assessments and other evaluation materials used to assess a child… are provided and administered in the child's native language or other mode of communication and in the form most likely to yield accurate information on what the child knows and can do academically, developmentally, and functionally, unless it is clearly not feasible to so provide or administer. (20 USC § 1414(c))

Determination of what type of assessment is most likely to produce valid and reliable information about a student who is a member of an ethnic and linguistic minority group is a multifaceted process that should involve considering the student's degree of acculturation, assessing linguistic proficiency, and reviewing empirical support for the use of potential measures (Acevedo-Polakovich et al., 2007). Research suggests that traditional cognitive measures underpredict academic achievement in CLD students, which violates an underlying assumption of the discrepancy model (Figueroa & Sassenrath, 1989). However, the National Association of School Psychologists (NASP) cautions against using interpreters to translate tests, which introduces error into the assessment, or relying solely upon nonverbal measures, which still reflect cultural bias and are often linguistically mediated (Blatchley & Lau, 2010). Nonverbal cognitive scores are also poorer predictors of academic achievement in CLD students than verbal or quantitative reasoning scores – a concerning finding given the frequency of their use with this population (Lakin, 2012). Most school psychologists see the assessment of students with limited English proficiency by bilingual examiners as ideal (Bainter & Tollefson, 2003). However, only about 11% of school psychologists are bilingual (Curtis et al.,

2008). Further, even bilingual psychologists often fail to review research when choosing appropriate measures for CLD students (O'Bryon & Rogers, 2010).

***RTI and CLD students*.** In sum, the traditional special education referral and assessment process presents numerous difficulties when it is used to identify SLD in CLD students. Scholars have argued that RTI models have the potential to remediate these difficulties and address the issue of CLD disproportionality in special education by reducing the number of students who are referred for special education under the false assumption that their academic difficulties are the result of learning disabilities (Klingner, Artiles, & Mendez Barletta, 2006; Xu & Drame, 2008). Indeed, studies of schools without well-established prereferral intervention systems have shown that struggling CLD students are often referred for evaluation based solely on teacher perceptions and with minimal to no use of early intervention strategies (Carrasquillo & Rodriguez, 1997; Klingner & Harry, 2006). By giving students numerous opportunities to respond to instruction that is culturally responsive, evidence-based, and increasingly intensified to meet their needs, practitioners ensure that the SLD classification is reserved for students who truly meet the definition by ruling out alternative explanations for school failure, including linguistic proficiency (Brown & Doolittle, 2008; Klingner et al., 2006).

Meeting the needs of CLD students within the context of RTI models of service delivery is an emerging topic within the field. Studies have documented the efficacy of comprehensive reading interventions for young CLD students. Kamps et al. (2007) implemented a three-tier RTI model for CLD students, with tier 2 consisting of small-group intervention in targeted areas of reading deficiency, and compared the gains made

by these students with those of CLD students at other schools who received only CLD instruction. Students educated within the RTI framework made significantly greater gains than their CLD instruction-only counterparts (Kamps et al., 2007). As part of another extensive research project, Vaughn, Linan-Thompson et al. (2006) and Vaughn, Mathes et al. (2006) screened first grade CLD students and randomly assigned those who were at risk of reading failure to participate in a targeted tier 2 intervention, or receive only their regular CLD instruction. The intervention was implemented in Spanish for some students (Vaughn, Linan-Thompson et al., 2006) and English for others (Vaughn, Mathes et al., 2006), and was matched to the primary language utilized in each school's CLD program. Students receiving the Spanish intervention made significant gains in fluency, phonological awareness, phonics, and comprehension only when assessed in Spanish (Vaughn, Linan-Thompson et al., 2006). Conversely, those receiving the English intervention showed the most gains in English, although these students also improved in Spanish phonological awareness and comprehension skills (Vaughn, Mathes et al., 2006). Across both studies, 50 of 53 students responded to their respective interventions, and gains were maintained through the end of second grade, with a majority of students who had received intervention no longer falling into the at-risk range (Linan-Thompson, Vaughn, Prater, & Cirino, 2006; Cirino et al., 2009).

Research has also explored specific issues in the screening and progress monitoring of CLD students as part of RTI practice. Oral reading fluency CBMs in both Spanish and English have been shown to correlate well with one another, particularly in lower grade levels, as well as successfully predict academic outcomes in CLD students

(Keller-Margulis, Payan, & Booth, 2012; Ramírez and Shapiro, 2007). In general, CLD students tend to score lower on these measures and make slower, more curvilinear growth across school years than students who are proficient in English (Al Otaiba et al., 2009; Logan & Petscher, 2010; Ramírez & Shapiro, 2007). CLD students may also be less likely to show within-school-year differences in growth rates on ORF measures (i.e., more growth in the fall than in the spring; Keller-Margulis et al., 2012). Importantly, research indicates that ORF growth rates can be used to reliably differentiate between CLD students who do and do not receive special education services for SLD (Al Otaiba et al., 2009). Less research has been completed on the use of maze CBMs, but as is the case with English-proficient students, these measures appear to show adequate, if less robust, psychometric properties than ORF measures (McMaster, Wayman, & Cao, 2006).

Even less is known about the utility of writing and math CBMs with CLD students. Campbell et al. (2013) and Espin et al. (2008) have studied the use of writing CBMs for progress monitoring in secondary school, and found them to have adequate reliability and validity. Interestingly, scores on writing CBMs may actually be more strongly related to state test scores for CLD students than non-CLD students (Espin et al., 2008). Even so, there is a clear need for further research on writing CBMs, particularly with younger CLD students. To date, no studies have examined the reliability and validity of math CBMs with this population.

Taken together, these findings imply that CLD students can and often do respond to intensive early intervention, even with the added challenges they face due to their linguistic proficiency status. For some, intervention can remediate difficulties that would

have otherwise caused them to be referred for special education. Moreover, reading interventions have been found to be successful regardless of whether they are delivered in Spanish or English – an important factor in integrating CLD students within a school's broader RTI framework. CLD students typically make slower progress than native English speakers (Al Otaiba et al., 2009), but by providing enough time for these students to grow, the RTI process can identify CLD children who persistently do not respond to academic interventions and may truly have learning disabilities (Brown & Doolittle, 2008; Kamps et al., 2007). Thus, even if RTI is not formally used to make a disability determination, the data it produces has great potential to inform the psychoeducational evaluation process for these students.

At the same time, not all of the research on the use of RTI models of service delivery with CLD students in schools has yielded positive results. In a longitudinal study of data from 142 school districts in California, Bouman (2010) demonstrated that implementation of RTI reduced the overall rate of children identified as learning disabled, but failed to change the disproportional representation of CLD students within this group. In a qualitative study of RTI practices with CLD students in a single school, Orosco and Klingner (2010) identified several major flaws, including tier 1 general education practices that emphasized mainstream cultural experiences, tier 2 instruction that falsely assumed students had good understanding of English phonemes, and tier 3 instruction that was unresponsive to student difficulties. Despite the fact that students were regularly progress monitored in oral reading fluency, this data was underutilized in determining whether or not to change or intensify interventions, as well as in making

eligibility determinations for children who were evaluated for special education services (Orosco & Klingner, 2010). Thus, to date, the evidence on RTI models and CLD students is not conclusive enough to support a call for widespread implementation of RTI models with all CLD students (Hernández Finch, 2012). There is a clear need for additional research on the role of RTI in meeting the needs of CLD students, from early intervention to special education eligibility decision-making.

The Present Study

The purpose of the present research is to explore how school psychologists integrate traditional assessment methods with RTI data in the psychoeducational evaluation process for learning disabled students. This study will analyze the type and amount of RTI data included in evaluations of students who have qualified for special education services under the SLD category, as well as explore how the presence or absence of such data affected the selection of traditional cognitive and academic measures and the determination of legal exclusions Given RTI's hypothesized value in meeting the needs of CLD students, this subgroup has been chosen for focused analysis.

This work builds upon pilot research (Cohen, Sullivan, & Caterino, 2009; Sullivan, & Cohen, 2012), in which initial psychoeducational evaluations for students with SLD were analyzed in terms of their integration of traditional assessment methods with RTI data. In their study of 42 psychoeducational evaluation reports, Sullivan and Cohen (2012) found that referencing RTI was uncommon, and that inclusion of actual RTI data was even rarer, with only 11% of their sample providing progress monitoring data to show how students responded to academic interventions. They also noted that

many reports referenced accommodations and modifications to the instructional environment, while only 7% described evidence-based academic interventions. The present study will allow for more in-depth study of these topics than has been done in the past, because it will consider the specific strategies, interventions, and types of RTI data named in reports. Moreover, to date, no published research study has compared the integration of RTI data and traditional assessment information between CLD and non-CLD students. The present study aims to fill that gap in the literature by considering differences between these two groups in the way RTI is incorporated into reports, as well as the standardized cognitive and achievement tests utilized. It will also provide much-needed information on how school psychologists rule out exclusionary factors for SLD in CLD students, particularly linguistic proficiency.

Data for the current study are drawn from a local school district. Since the start of the 2005-2006 school year, all school psychologists in the district have been trained in response to intervention frameworks and how to use data, including universal screening and progress monitoring scores, as a component of the learning disability identification process. This district is permitted to use RTI in determining SLD eligibility by the state department of education. It has also been recognized as a "best practice" district by a major nonprofit organization due to the strength of its RTI model. According to documentation provided by the district, the model involves three tiers of support, beginning with a quality general education curriculum. It is assumed that between 16 and 20% of students will require targeted, tier 2 intervention, and another 3 to 5% will require intensive, tier 3 intervention. All educational practices, from the core curriculum to the

most intensive intervention techniques utilized, are evidence-based and implemented with fidelity and consistency.

Psychoeducational reports have been selected as a data source because they represent a consistent and crucial product of the multidisciplinary evaluation process. The necessary components of a psychoeducational evaluation are outlined in IDEIA (2004) and adopted by individual states, generally with few additions or modifications. Thus, analyzing psychoeducational reports presents a way not only to understand how practitioners approach the process of determining whether or not a student qualifies for and is in need of special education, but also to gauge practitioners' compliance with legal standards.

Yet, despite the utility of psychoeducational reports, relatively few published studies have employed them as a data source. Some of this work has addressed the clarity and utility of psychoeducational evaluations written in private practice or pediatric hospital settings (e.g., Cornwall, 1990; Harvey, 1997). Although reports written in these settings may have much in common with reports written by school-based practitioners, they are not subject to the same legal requirements under IDEIA (2004). Much of the research on school reports has focused on the way they are written, addressing topics such as readability and social validation (Weddig, 1984; Wiener, 1987), organization and structure (Pryzwansky & Hanania, 1986), the nature of recommendations (Mallin, Schellenberg, & Smith, 2012), the incorporation of graphs (Miller & Watkins, 2010), and the way that reports affect teacher perceptions of students (Schwartz & Wilkinson, 1987). In many of these studies, participants were asked to read fictitious reports created

specifically to address the variables in question. Thus, the potential for actual psychoeducational reports written by school psychologists to provide data on evaluation practices and compliance with special education law has been largely overlooked.

The present study will thus expand on Sullivan and Cohen's (2012) work and fill an existing area of deficit in the literature. To date, the only published study with a similar method and focus was completed by Wilkinson et al. (2006), who analyzed 21 educational files, (including psychoeducational reports) of CLD students who had been classified as learning disabled in reading. In that study, three doctoral-level bilingual special education faculty members reviewed these records in terms of whether or not they would, as professionals, have classified the students as SLD, what data was or was not considered in the actual school team's decision, and whether or not the exclusionary clauses were truly considered. Wilkinson et al.'s work indicated that evaluations often fail to comply with legal standards for eligibility determinations, particularly when it comes to ruling out inadequate instruction, linguistic proficiency, and environmental, economic, and cultural disadvantage as factors in poor academic performance. The scope of the present study will allow for analysis of the level of compliance with special education law in psychoeducational evaluations of students with learning disabilities, but with a sample that includes both CLD and non-CLD students and is much larger than what has been used in previous research.

Research question 1: How does the inclusion of RTI data influence the psychoeducational evaluation process? One of the primary research questions of this study concerns the degree to which RTI data is incorporated into evaluations across the

full sample of reports completed on both CLD and non-CLD students. This study will provide in-depth information on how RTI influences the psychoeducational evaluation process, including specific interventions and strategies (including accommodations) documented in reports, as well as the presence, type, and amount of RTI data included.

Hypothesis 1a. Given that attention to RTI models of service delivery and the use of RTI data in special education evaluations has increased steadily over time, it is anticipated that, within this study's sample, reports completed within the last two years will be more likely to reference RTI than reports completed within the first two years. As RTI is a broad term, this may be operationalized in terms of the inclusion of statements that prereferral strategies and/or interventions were tried, as well as information on the specific strategies and interventions attempted.

Hypothesis 1b. However, it is also anticipated that, among reports stating that prereferral strategies and interventions had been attempted, only a small percentage will include concrete information on the specific research-based interventions provided, and that many will list strategies (including accommodations) rather than true interventions. Moreover, within this study's sample, reports from the most recent two years will be more likely to reference research-based interventions than reports from the previous two years.

Hypothesis 1c. Similarly, it is expected that only a small percentage of reports referencing prereferral strategies and interventions will include actual RTI data, defined as universal screening scores and data on the students' progress while receiving interventions.

Hypothesis 1d. It is further expected that reports completed on elementary school students will incorporate a greater degree of information on the RTI process than reports completed on secondary (middle and high school) students. This will be operationalized in terms of the number of data points included in each report, the number of interventions cited, and the length, frequency, and timing of interventions.

Hypothesis 1e. Moreover, it is hypothesized that significantly more reports will contain RTI data in reading than any other academic area.

Hypothesis 1f. Finally, it is expected that reports referencing the RTI process (including citing prereferral strategies and interventions, as well as RTI data) will be likely to use fewer traditional assessments than reports without this information.

Research question 2*: *How does the psychoeducational evaluation process differ for CLD students, particularly with regard to the inclusion of RTI data?

Hypothesis 2a. With regard to traditional assessment alone, it is anticipated that there will be greater variability in the measures utilized in evaluations completed for CLD students than evaluations completed on non-CLD students. This is expected to result in a higher mean number of assessments (overall and within each domain) per report, as well as in a broader range of tests across the subsample.

Hypothesis 2b. It is also anticipated that cognitive assessment of CLD students is more likely to rely on nonverbal measures of ability than cognitive assessment of non-CLD students. Nonverbal measures will be considered assessments that are designated as such in their publication titles and descriptions. If a broad cognitive measure has a

separate index designated as “nonverbal” and it is used in lieu of the rest of the measure, it may also be considered a distinct nonverbal assessment for the purposes of this study.

Hypothesis 2c. Within the CLD subsample alone, nonverbal measures are expected to be the most commonly used tests of cognition, followed by traditional IQ tests given in English, and lastly by IQ tests administered in Spanish. Academic assessments in English are expected to be more commonly used than academic assessments in Spanish.

Hypothesis 2d. It is anticipated that evaluations of CLD students will be more likely than evaluations of non-CLD students to go beyond simple measures of IQ and achievement and utilize additional pieces of evidence to prove that the student’s failure to learn is not due to exclusionary factors. These sources of information may include data on student responses to targeted intervention, but also may involve data on their progress in the general curriculum (e.g., classroom benchmark scores).

Hypothesis 2e. Within the CLD subsample alone, the specific methods by which school psychologists rule out linguistic proficiency as the chief factor in learning difficulties are expected to be highly variable. Information contributed by parents (including home language survey results),educational history, existing state proficiency test data, and performance on individually-administered proficiency measures are anticipated to emerge as tools used in this process. Assessment methods are also expected to be related to the psychologist’s bilingual status.

CHAPTER 2

METHOD

Participants

The final sample includes 212 students between the ages of 5 and 18 who were the subjects of initial psychoeducational evaluations during the 2009-2010, 2010-2011, 2011-2012, and 2012-2013 school years. Psychoeducational evaluation reports were sampled from a large, unified school district in the southwestern United States that serves approximately 26,000 students from preschool through 12th grade. Students in the district speak more than 60 languages and come from a variety of socioeconomic, racial, and cultural backgrounds. The most recent available census data indicates that about 67% of youth within the district's boundaries are White, 2% are Black, 16% are Hispanic, 4% are Asian, 6% are of other ethnic backgrounds, and 5% are of mixed race (U.S. Department of Education NCES, 2013). According to the state department of education website, 27% of students in the district were eligible for free or reduced lunch in 2012; however, figures by school varied widely, from 3% to 86% (ADE, 2012a). Additionally, about 4% of students in the district were classified as having limited English proficiency in 2011, the most recent year for which data is available; this figure has declined since 2008, when it was roughly 7% (New America Foundation, 2013). Since 2000, all schools in the state have been required by law to use English-only instruction with all students, including those with limited English proficiency (Lawton, 2012).

Procedure

Copies of all initial psychoeducational evaluations resulting in eligibility for SLD across the most recent four school years (2009-2010, 2010-2011, 2011-2012, 2012-2013) were obtained from the school district. Access to school psychologists' annual evaluation logs from the school years selected for analysis was provided to the researcher. In order to qualify for selection, reports had to have specifically determined that the student qualified for special education services under the SLD category in one or more of the eight qualifying areas. In keeping with the present study's focus on how assessment methods were integrated for students with identified learning disabilities, reports completed on students who ultimately did not qualify for special education were not included. Acceptance reports of previous evaluations from outside the district were also excluded because they did not include any assessments conducted by district psychologists. Reevaluations, including those completed on students who previously qualified for preschool special education services, were excluded from the sample, because the focus of the study was on the incorporation of RTI data in making initial eligibility decisions. An exception to this rule was extended to reports on students who were previously classified as only having speech-language impairment (SLI). Such reports were included in the present sample, because they represented the first time students had been assessed cognitively and academically.

A total of 260 reports were initially identified for inclusion based on school psychologists' logs across the four school years studied. Of these, 212 reports were recovered from student files and found to meet all necessary criteria. Forty-eight reports

were not included because they either could not be located or were found to not actually meet the inclusion requirements; for example; reports were excluded if they were reevaluations or acceptance reports despite having been listed as initial evaluations on the logs.

Reports were written by 25 certified school psychologists, four of whom were bilingual; all bilingual psychologists spoke Spanish as their second language. Bilingual psychologists were assigned to work at specific schools within the district, but were also available to perform evaluations on students at other schools if needed. Five school psychologists were licensed by the state board of psychologist examiners, and eighteen held doctoral level degrees. All evaluators were certified as school psychologists by the state department of education.

All psychoeducational reports meeting the inclusion criteria were blinded and assigned an ID number for the purposes of this study, and no identifying information about students, teachers, parents, or schools appeared on the copies used for coding or the coding protocol. Ten percent of the reports in the final sample were randomly selected for review by a second rater in order to obtain an interrater reliability estimate.

Thirty-two reports were completed on students who had ELL status noted on their initial IEPs that were written immediately after they were evaluated and qualified for special education services under the SLD category. An additional 18 reports were completed on students who, according to their psychoeducational evaluations, had been exited from ELL services, but still had a home language other than English listed on their IEPs. In keeping with the definition of a CLD student as one whose background differs

culturally and linguistically from that of the mainstream population (Watson et al., 2007), these students were also included in the CLD subgroup for analysis, bringing the total to 50 evaluations (24% of the total sample).

Instrument

The coding protocol for this study was inspired by the work of Sullivan and Cohen (2012). The original protocol was developed in accordance with special education law in the state in which the study was conducted (ADE, 2009a), and thus closely followed federal assessment and eligibility guidelines. For the purposes of the present study, the existing protocol was modified and expanded significantly; this version appears in the Appendix. Although most items regarding compliance with the law were retained due to the fact that special education law had not measurably changed since 2009, several key changes were made for clarity and to allow for collection of a wider range of data.

The coding protocol for the present study addressed seven domains: (1) basic information, (2) review of existing data, (3) overall disability rule-outs, (4) assessment tools, (5) statement of SLD and SLD rule-outs, (6) RTI, and (7) language. The first five domains were completed for all reports reviewed. Completion of the final two was contingent on the inclusion of RTI information in the evaluation and the description of the student's language status.

Basic information. Basic information included demographic information on the student, as well as the cognitive, academic, social-emotional, and other assessments utilized. The degrees and credentials (e.g., licensure as a psychologist) and bilingual

status of the practitioner who completed the evaluation were also addressed. This section included eleven items. Items addressing the race/ethnicity and gender of the student, the school year in which the report was written, and psychologist attributes, were strictly categorical in nature. Items addressing assessments were categorical as well, with potential responses based on consultation with the school district's lead psychologist regarding common measures used by practitioners in the district; however, these items also included write-in spaces for any specific assessments not captured by the categories. Items addressing age and grade level of the student were open-ended. The item about school year was added to allow for comparisons to be made in the overall level of integration of RTI data across the district over the four school years.

Review of existing data. This section addressed whether or not the report included the information legally required to be present in the review of background data. As dictated by Arizona Revised Statutes (ARS) §300.305, this consisted of information provided by parents, review of current classroom performance and local/state assessments, and observations from teachers and related service providers. In addition, it addressed the requirement in ARS §300.304 that functional and developmental information be reviewed, as well as the requirement that the evaluation process include a classroom-based observation of the student. Eight items were included in this section, and all were dichotomous (coded yes or no depending on the presence or absence of the information in the report).

Overall disability rule-outs. This section addressed the rule-outs for disability status discussed in ARS §300.305 and ARS §300.306. These were lack of appropriate

instruction in reading, lack of appropriate instruction in math, and limited English proficiency. A total of eleven items were included in this area, all of which were dichotomous. Items targeted whether or not the rule-out was addressed in the evaluation (including if it was treated as a disclaimer, meaning a brief statement without explanation). For the instructional areas, items addressed the presence of evidence that appropriate general education instruction had been provided (e.g., information on curriculum and/or interventions used), as well as the presence of data to support progress monitoring in the general curriculum (e.g., CBM data or classroom benchmark assessment scores). For the linguistic rule-out, items addressed whether language proficiency was mentioned as a disclaimer or more fully described.

Assessment tools. This area focused on legal requirements pertaining to the selection and use of assessment tools, as dictated by ARS§300.304. Items addressed whether or not a variety of assessment instruments (more than one) were used, and whether or not evaluation procedures were selected and administered so as not to be racially or culturally discriminatory. Information on the variety of assessment tools and strategies) was captured in one item.. In total, this section contained five dichotomously-scored items.

Statement of SLD and SLD rule-outs. This section included items relating to the formal statement of whether or not a student had SLD, as discussed in ARS§300.311. Items addressed how the determination was made (discrepancy versus RTI), as well as other required information (e.g., incorporation of the child's strengths and weaknesses). SLD rule-outs were also discussed as per ARS§300.309. Importantly, the rule-outs

regarding cultural factors and environmental/economic disadvantage were included in this section because, in the law, they are SLD-specific rule-outs rather than general disability rule-outs. Rule-outs for visual, hearing, and motor disabilities were addressed in separate items to allow for specificity in coding. Separate items also addressed the inclusion of information about whether or not a student was meeting grade level standards and making progress with regard to those standards. In addition, an item was included in this section regarding what SLD subcategories a child qualified under, with choices reflecting federal eligibility categories outlined in IDEIA (2004). This was added to allow for analysis of differences in how RTI data is used across subject areas (e.g., for students with difficulties in reading versus students with difficulties in mathematics).

In total, there were 31 items in this section, 30 of which were dichotomous. The final item regarding SLD subcategorization was categorical, and included choices for all eight potential areas of SLD addressed in ARS §300.309: Basic Reading, Reading Fluency, Reading Comprehension, Math Calculation, Written Expression, Oral Expression, and Listening Comprehension).

RTI. This portion of the coding scheme focused on how data addressing students' ability to respond to intervention was incorporated into reports. Importantly, RTI did not have to be formally used as a way of determining eligibility – the mention of RTI and/or the presence of RTI data in the report was sufficient. Items addressing the documentation of strategies that had been tried and determined to be successful or unsuccessful were included. An item was included to address whether specific interventions were listed, and if so, what specific interventions were included in the report and how long they were

implemented. In order to be considered an intervention, a method listed in a report must have met the broad criteria described by the state department of education (ADE, 2012b). ADE (2012b) defines an intervention as:

> The directing of instruction in the area(s) of concern that is in addition to regular classroom instruction. Interventions are designed to meet the identified needs of an individual and are monitored on regular and frequent basis. Changes in instruction, for the student in the area of learning difficulty, are designed to improve learning and to achieve adequate progress. (p. 4)

Though this definition is broad, it is also limited in that it may not include some types of strategies implemented for specific students to address learning problems. For example, it would not include accommodations, which are defined as supports that help facilitate access to the general education curriculum (e.g., extended time, use of a separate setting for exams, reading test materials aloud; ADE, 2012b). Given that instructional strategies, including accommodations, are often part of the prereferral process in addition to formal interventions (NCRTI, 2010), a section was added to address whether additional strategies (including accommodations) were discussed and, if so, what they were. The definition of an accommodation outlined by ADE was used as the basis for coding in this area.

These items focused on the presence of can't do/won't do assessment information, universal screening data, and progress monitoring data for students receiving intervention were also added. These areas were included because of their status as key components of RTI models of service delivery (Burns & Coolong-Chaffin, 2006; VanDerHeyden &

Witt, 2008). For the latter two areas, the coding scheme also included questions about the type of data presented (e.g., oral reading fluency, math calculation), as well as the number of data points provided in the report. Additionally, an item addressing whether or not a specific decision rule was referenced to determine responsiveness was added given the wealth of literature on this topic (e.g., Ardoin et al., 2013; Compton et al., 2006).

In total, the RTI section included 19 items, 11 of which were dichotomous. The two items addressing the type of universal screening and progress monitoring data included were categorical in nature; each included four choices (oral reading fluency words per minute, mazes, math calculation digits, or other). These potential responses were derived from documentation provided by the district on the types of screening and progress monitoring measures they used. The two items regarding the number of data points allowed for numerical responses. Finally, the remaining items on specific accommodations used, specific interventions used, the length of time interventions were implemented, and who implemented each intervention were open-ended.

Language. The final section on language was not to be completed if the student had not been identified as CLD during the sampling process, and if the student's home, dominant, and first languages were all identified as English in the report. Items in this section addressed the identification of the child's home, dominant, first, and most proficient languages, the inclusion of the length of time the student had lived in the United States, the child's educational history, and the discussion of the impact of language proficiency on educational progress. This area also addressed the requirement

outlined in ARS§300.304 that evaluation procedures be conducted in the method of communication most likely to yield accurate information about the child's functioning. Items were included about language proficiency and how it was determined. Overall, a total of twelve items were part of this final section of the coding protocol. All items were dichotomously scored; however, items addressing the inclusion of information about the home language, student's primary language, student's most proficient language, student's first language, and inclusion of proficiency test information also allowed for open-ended responses.

It is important to note that this section was created not only with consideration of federal special education law, but also with regard to state policies on culturally and linguistically diverse students in general. In Arizona, a Home Language Survey is provided to parents and used to determine children's placement in ELL classes. Prior to 2009, this survey included three questions that asked about the child's home language, the language most often spoken by the child, and the child's first language – a yes to any of the three led to the student's legal classification as ELL (Goldenberg & Rutherford-Quach, 2012). However, in July 1, 2009, the survey was changed to include only one question – "what is the primary language of the student?" (ADE, 2009b). Notably, research suggested that use of a single-question survey resulted in many diverse learners not receiving the services to which they were entitled (Goldenberg & Rutherford-Quach, 2012). On April 1, 2011, after a settlement with the U.S. Department of Education's Office for Civil Rights and the U.S. Department of Justice, Civil Rights Division, a new survey went into effect that once again included the three questions (ADE, 2011; U.S.

Department of Education, 2011). Because the sample for the present study includes reports written when both the single-question and three-question language surveys were in effect, both ways of describing proficiency could appear in psychoeducational reports. Thus, the coding protocol was written to cover all three potential home language questions.

The study was approved as exempt pursuant to federal regulations by the Arizona State University (ASU) Institutional Review Board (IRB); the exemption letter is reproduced in the Appendix.

CHAPTER 3

RESULTS

Data Procedures

Power analysis. Power analysis completed using the G*Power software program (Faul, Erdfelder, Buchner, & Lang, 2009) indicated that the total sample size of the study ($N = 212$) exceeded the minimum required to achieve Cohen's (1992) recommended power level of .80 for the expected tests of nominal data involved in the study ($N = 143$). Additional power analysis indicated that the size of the CLD subgroup ($n = 50$) was greater than the minimum required to achieve a power level of .80 for the planned tests of numerical data that were proposed ($n = 39$).

Statistical procedures. Data analysis began on a descriptive level. First, interrater reliability was calculated with 5% of the reports using Cohen's kappa (Cohen, 1960); reports to be coded by the second rater were selected at random using a random number generator. The rate of agreement between raters was 88% ($\kappa = .82$), which exceeds the minimum recommended rate of 80%.

A majority of items in the coding scheme were dichotomously scored and concerned whether or not the reports contain specific elements. For each of these items, the percentage of reports including that element was calculated. In addition, all items that address the presence or absence of a state-required element in psychoeducational evaluations may be considered a subset that determines whether or not the report is fully in compliance with special education legislation. After initial analysis, these compliance-based items were analyzed separately to determine an overall proportion of reports that

did and did not adhere to legal criteria, as well as the primary reasons why noncompliant reports failed to meet this standard.

Additional items were considered categorical or open-ended in nature. Items addressing the age, race/ethnicity, grade, and SLD subclassification of the student had a limited number of potential responses, and thus the frequency of each potential response was easily calculated. However, items focusing on strategies and interventions utilized were more open-ended. For these items, all verbatim responses were initially recorded and then carefully examined to group together responses that are similar enough to be considered identical (e.g., different editions of the same instrument). Data was then generated on the frequency of each response. The same process was used to address responses to the items about instruments utilized and degrees/credentials of the psychologist, which, although categorical, contained "Other" categories for unlisted responses. Finally, the items on the coding protocol that dealt with the number of data points provided in universal screening and progress monitoring data, as well as the item about how long each intervention was implemented, yielded numerical responses.

More complex analyses followed the generation of initial descriptive statistics. A major goal of this study was to investigate the differences between evaluations of CLD students and evaluations of native English speakers. Thus, after the initial descriptive analysis, nonparametric tests were used to determine whether or not there were significant differences between these two groups in the nominal areas addressed by the coding protocol. In accordance with the recommendations of McDonald (2009), chi-squared or Fisher's exact tests were selected based on cell size, with the presence of cell

sizes of five or fewer necessitating use of Fisher's exact test. Nonparametric tests were also used for categorical analysis across years. Within-case categorical analysis (e.g., comparing proportions of reading and math data) was completed using McNemar tests. Finally, for items on the coding scheme that yielded numerical instead of categorical data, Mann-Whitney tests were used because the data for these items were not normally distributed and therefore violated the assumptions of traditional t-tests.

Empirically-based interventions. As the hypotheses for the present study addressed the presence or absence of empirically-based interventions in reports, a procedure for determining which interventions meet this criterion was performed. In 2002, the U.S. Department of Education's Institute for Education Science (IES) established the What Works Clearinghouse (WWC; 2014), an initiative that analyzes the extant literature on intervention programs used in schools and reports on the strength of the body of research to support the efficacy of these interventions. WWC (2014) classifies interventions as having Negative, Potentially Negative, No Discernible, Mixed, Potentially Positive, or Positive effects. For the purposes of this study, interventions classified by WWC as having Potentially Positive or Positive effects will be considered empirically-based. In many cases, WWC produces several reports on one intervention program, each of which discusses its impact on a particular area of academics. For interventions for which this is the case, the presence of a WWC report indicating Potentially Positive or Positive effects in at least one domain for elementary or secondary students was accepted as sufficient to designate the intervention as empirically-based. An additional widely used resource for disseminating information about empirically-

based interventions for reading is the Florida Center for Reading Research (FCRR, 2014). Like WWC, FCRR (2014) analyzes research on intervention programs and produces reports summarizing the strength of this research. In addition, FCRR produces its own research-based materials for teachers. Intervention programs supported by FCRR reports, or methods that are part of FCRR materials, were also considered empirically-based.

Sample Characteristics

The characteristics of the final sample by demographic variable are displayed in Table 1. Approximately 60% of the sample was composed of reports completed on male students, while 40% was composed of reports on female students. The mean age of students in the sample was 9.52 years. Most students were in 2nd grade (17%), 3rd grade (14%), 1st grade (13%), or 4th grade (12%) upon being evaluated. There was a significant difference between the racial/ethnic distribution of the sample and that of the district's overall population, $\chi^2(5, 157) = 26.96$, $p < .001$, with White students underrepresented in the sample and Hispanic students slightly overrepresented. The distribution of reports across school years was uneven, with 76 (36%) from the 2009-2010 school year, 58 (27%) from the 2010-2011 school year, 35 (17%) from the 2011-2012 school year, and 43 (20%) from the 2012-2013 school year. This difference was statistically significant, $\chi^2(3, 212) = 18.45$, $p < .001$. This may be attributable to differences in the number of school psychologists in the district who submitted logs, rather than actual differences in the number of evaluations taking place each year. However, it is also important to note that the district documented a decline in the number of students placed in special education under the SLD category across the years studied.

Table 1

Demographic Composition of the Sample

Demographic Variable	Full Sample (%)	Non-CLD Students (%)	CLD Students (%)
CLD status			
CLD	162 (76%)	--	--
Non-CLD	50 (24%)	--	--
Gender			
Male	128 (60%)	96 (59%)	32 (64%)
Female	84 (40%)	66 (41%)	18 (36%)
Ethnicity			
White	87 (41%)	85 (53%)	2 (4%)
Asian	10 (5%)	7 (4%)	3 (6%)
Black	4 (2%)	4 (3%)	0 (0%)
Hispanic	46 (22%)	14 (9%)	32 (64%)
American Indian	5 (2%)	4 (3%)	1 (2%)
Mixed	5 (2%)	5 (3%)	0 (0%)
Unspecified	55 (26%)	43 (27%)	12 (24%)
Grade level			
Kindergarten	17 (8%)	14 (7%)	3 (6%)
1st grade	27 (13%)	24 (15%)	3 (6%)
2nd grade	35 (17%)	29 (18%)	6 (12%)
3rd grade	29 (14%)	21 (13%)	8 (16%)
4th grade	26 (12%)	18 (11%)	8 (16%)
5th grade	20 (9%)	16 (10%)	4 (8%)
6th grade	14 (7%)	9 (6%)	5 (10%)
7th grade	8 (4%)	6 (4%)	2 (4%)
8th grade	17 (8%)	12 (7%)	5 (10%)
9th grade	6 (3%)	5 (3%)	1 (2%)
10th grade	5 (2%)	3 (2%)	2 (4%)
11th grade	3 (1%)	2 (1%)	1 (2%)
12th grade	2 (1%)	1 (1%)	1 (2%)
Unspecified	3 (1%)	2 (1%)	1 (2%)
Degree of psychologist			
MA	79 (37%)	63 (39%)	16 (32%)
MS	33 (16%)	27 (17%)	6 (12%)
PhD	86 (41%)	63 (39%)	23 (46%)
EdD	3 (1%)	1 (1%)	2 (4%)
PsyD	11 (5%)	8 (5%)	3 (6%)
Licensure status of psychologist			
No	182 (86%)	137 (85%)	45 (90%)
Yes	30 (14%)	25 (15%)	5 (10%)
Bilingual status of psychologist			
No	188 (89%)	146 (90%)	41 (84%)
Yes	24 (11%)	16 (10%)	8 (16%)

Note. Data are reported for the full sample (N = 212), as well as for the non-CLD (n = 162) and CLD (n = 50) subgroups.

A majority of reports (53%) were completed by masters-level clinicians. Licensure status was uncommon, with only 14% of reports completed by licensed practitioners. Likewise, bilingual status was somewhat rare, with 11% of reports done by bilingual school psychologists; all were proficient in English and Spanish. This proportion was somewhat greater within the CLD subgroup of reports (16%) than the non-CLD subgroup (10%).

General Descriptive Data

Descriptive data on the use of various assessment tools and methods in the identification of SLD was generated. The distribution of cognitive assessments present in reports is displayed in Table 2. Five reports in the sample did not include a cognitive measure administered as part of the psychoeducational evaluation process, but rather referenced cognitive testing completed by private practitioners or previous evaluators; however, all of these reports included new academic testing. The most commonly used cognitive assessments were, in order, the Wechsler Intelligence Scales for Children, 4th Edition (WISC-IV; 54%), Kaufman Assessment Battery for Children (KABC-2; 23%), and KABC-2 Nonverbal Index only (10%). This was true regardless of CLD status. However, the proportion of reports including the WISC-IV was about twice as large for non-CLD students (61%) than for CLD students (30%).Within the CLD subsample, the full KABC-2 and WISC-IV were present in an equal proportion of reports (30%). Also within this subgroup, some evaluators specified that they had elected to use a nonverbal index score as the best estimate of the child's ability, despite having included data from a full cognitive measure in their reports. The WISC-IV Perceptual Reasoning Index (PRI)

was used for this purpose in one report, the Stanford-Binet 5th Edition (SB-5) Quantitative Reasoning scale was used in one report, and the Reynolds Intellectual Assessment Scale (RIAS) Nonverbal Intelligence Index (NIX) was used in one report. The WISC-IV PRI, SB-5 Quantitative Reasoning scale, and RIAS NIX are not depicted in Table 2 because they were not administered in isolation. Additionally, of the 15 reports on CLD students that included the full KABC-2, eight stated that the Nonverbal Index was used as the best estimate of the child's intelligence.

Table 2

Use of Cognitive Assessments in Evaluations of CLD and Non-CLD Students with SLD

Assessment Name	Full Sample (%)	Non-CLD Students (%)	CLD Students (%)
Wechsler Intelligence Scales for Children, 4th Edition (WISC-IV)	113 (54%)	98 (61%)	15 (30%)
Kaufman Assessment Battery for Children, 2nd Edition (KABC-2)	49 (23%)	34 (21%)	15 (30%)
KABC-2 Nonverbal Index (NVI) only	21 (10%)	14 (9%)	7 (14%)
Kaufman Brief Intelligence Test, 2nd Edition (KBIT-2)	12 (6%)	8 (5%)	4 (8%)
Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III COG)	11 (5%)	10 (6%)	1 (2%)
Universal Nonverbal Intelligence Test (UNIT)	8 (4%)	2 (1%)	6 (12%)
Reynolds Intellectual Assessment Scales (RIAS)	6 (3%)	4 (3%)	2 (4%)
Stanford-Binet Intelligence Scales, 5th Edition (SB-5)	6 (3%)	5 (3%)	1 (2%)
Test of Nonverbal Intelligence, 3rd Edition (TONI-3)	5 (2%)	3 (2%)	2 (4%)
No cognitive measure administered	5 (2%)	4 (3%)	1 (2%)
Differential Ability Scales, 2nd Edition (DAS-2)	3 (1%)	3 (2%)	0 (0%)
Comprehensive Test of Nonverbal Intelligence (CTONI)	3 (1%)	0 (0%)	3 (6%)
Wechsler Adult Intelligence Scales, 4th Edition (WAIS-IV)	1 (1%)	1 (1%)	0 (0%)
Wechsler Nonverbal Scale of Ability (WNV)	1 (1%)	0 (0%)	1 (2%)
Wechsler Preschool and Primary Scale of Intelligence, 4th Edition (WPPSI-IV)	1 (1%)	1 (1%)	0 (0%)
KABC-2 NVI administered in Spanish	1 (1%)	0 (0%)	1 (2%)
WJ-COG administered in Spanish	1 (1%)	0 (0%)	1 (2%)
Draw-a-Person Test (DAP)	1 (1%)	1 (1%)	0 (0%)
Total number of cognitive assessments identified	17	13	13

Note. Each percentage reflects the proportion of reports in the full sample (N = 212) including that assessment. Data are further reported for the proportion of non-CLD reports (n = 162) and CLD reports (n = 50) that included each assessment. The most recent version of each assessment used is named, but earlier editions were also included for ease of analysis. The KABC-2 Nonverbal Index was the only nonverbal portion of a full cognitive test that was administered in isolation.

The distributions of academic, social/emotional/behavioral, and other assessments are outlined in Tables 3, 4, and 5, respectively. Academic assessment data were similarly distributed across the subgroups, with the Woodcock-Johnson III Tests of Academic Achievement (WJ-III ACH; 66%) and the Wechsler Individual Achievement Test, 3rd Edition (WIAT-III; 38%) being the most commonly used for both groups. However, within the CLD subgroup, the Batería III Woodcock-Muñoz Achievement Battery (8%) was the third most commonly used assessment, whereas several other additional achievement measures were often used with non-CLD students. Across the full sample, fewer than half of all reports (42%) did not include a specific social, emotional, or behavioral measure. Beyond that, the most common measure included in reports of both CLD and non-CLD students was the Behavior Assessment System for Children (BASC-2; 39%), followed by an informal, unstandardized behavioral/social screening form (26%) produced by the school district that was completed by parents. A majority of reports did not include additional types of measures. However, for reports that did, it was most common to include the Beery-Buktenica Test of Visual-Motor Integration (20% of the full sample).

Table 3

Use of Academic Assessments in Evaluations of CLD and Non-CLD Students with SLD

Assessment Name	Full Sample (%)	Non-CLD Students (%)	CLD Students (%)
Woodcock-Johnson III Tests of Academic Achievement (WJ-III)	140 (66%)	103 (64%)	37 (74%)
Wechsler Individual Achievement Test, 3rd Edition (WIAT-3)	81 (38%)	64 (40%)	17 (34%)
Test of Early Reading Ability, 3rd Edition (TERA-3)	13 (6%)	11 (7%)	2 (4%)
Test of Early Mathematics Ability, 3rd Edition (TEMA-3)	12 (6%)	10 (6%)	2 (4%)
Kaufman Test of Educational Achievement, 2nd Edition (KTEA-2)	10 (5%)	8 (5%)	2 (4%)
Test of Early Writing Ability, 3rd Edition (TEWL-3)	7 (3%)	5 (3%)	2 (4%)
Test of Written Language, 3rd Edition (TOWL-3)	6 (3%)	6 (4%)	0 (0%)
Basic Reading Inventory (BRI)	4 (2%)	3 (2%)	1 (2%)
Batería III Woodcock-Muñoz Achievement Battery	4 (2%)	0 (0%)	4 (8%)
KeyMath3 Diagnostic Assessment	3 (1%)	3 (2%)	0 (0%)
No academic assessment administered	2 (1%)	1 (1%)	1 (2%)
Curriculum-based measurement writing sample	1 (1%)	0 (0%)	1 (2%)
Test of Reading Comprehension, 3rd Edition (TORC-3)	1 (1%)	1 (1%)	0 (0%)
Young Children's Achievement Test (YCAT)	1 (1%)	0 (0%)	1 (2%)
Total number of academic assessments identified	13	10	10

Note. Each percentage reflects the proportion of reports in the full sample (N = 212) including that assessment. Data are further reported for the proportion of non-CLD reports (n = 162) and CLD reports (n = 50) that included each assessment. The most recent version of each assessment used is named, but earlier editions were also included for ease of analysis. In one report, a curriculum-based measurement tool in writing was discussed, but treated as a formal assessment in conjunction with other academic tests, rather than as a universal screening or progress monitoring tool; thus it is included here.

Table 4

Use of Social, Emotional, and Behavioral Assessments in Evaluations of CLD and Non-CLD Students with SLD

Assessment Name	Full Sample (%)	Non-CLD Students (%)	CLD Students (%)
No social, emotional, or behavioral measure administered	88 (42%)	63 (39%)	25 (50%)
Behavior Assessment System for Children, 2nd Edition (BASC-2)	82 (39%)	64 (40%)	18 (36%)
Informal behavioral/social screening form	56 (26%)	42 (26%)	14 (28%)
Conners 3rd Edition (Conners-3)	17 (8%)	15 (9%)	2 (4%)
Student interview	15 (7%)	14 (9%)	1 (2%)
Behavior Rating Inventory of Executive Function (BRIEF)	3 (1%)	2 (1%)	1 (2%)
Clinical Assessment of Behavior (CAB)	3 (1%)	3 (2%)	0 (0%)
Conners Comprehensive Behavior Rating Scales (CBRS)	3 (1%)	3 (2%)	0 (0%)
Projective drawings	3 (1%)	3 (2%)	0 (0%)
Behavior Evaluation Scale, 2nd Edition (BES-2)	2 (1%)	1 (1%)	1 (2%)
Emotional Disturbance Decision Tree (EDDT)	2 (1%)	2 (1%)	0 (0%)
Asperger Syndrome Diagnostic Scale (ASDS)	1 (1%)	1 (1%)	0 (0%)
Attention Deficit Disorders Evaluation Scale, 3rd Edition (ADDES-3)	1 (1%)	1 (1%)	0 (0%)
Autism Diagnostic Observation Schedule (ADOS)	1 (1%)	1 (1%)	0 (0%)
Autism Spectrum Rating Scales (ASRS)	1 (1%)	1 (1%)	0 (0%)
Children's Depression Inventory, 2nd Edition (CDI-2)	1 (1%)	1 (1%)	0 (0%)
Developmental Profile, 3rd Edition (DP-3)	1 (1%)	1 (1%)	0 (0%)
Functional behavioral assessment	1 (1%)	0 (0%)	1 (2%)
Social Responsiveness Scale (SRS)	1 (1%)	1 (1%)	0 (0%)
Spence Children's Anxiety Scale (SCAS)	1 (1%)	1 (1%)	0 (0%)
Total number of social, emotional, and behavioral assessments identified	19	18	7

Note. Each percentage reflects the proportion of reports in the full sample (N = 212) including that assessment. Data are further reported for the proportion of non-CLD reports (n = 162) and CLD reports (n = 50) that included each assessment. The most recent version of each assessment used is named, but earlier editions were also included for ease of analysis.

Table 5

Use of Other Assessments in Evaluations of CLD and Non-CLD Students with SLD

Assessment Name	Full Sample (%)	Non-CLD Students (%)	CLD Students (%)
No additional measures given	153 (72%)	115 (71%)	38 (76%)
Beery-Buktenica Test of Visual-Motor Integration (Beery VMI)	43 (20%)	36 (22%)	7 (14%)
Comprehensive Test of Phonological Processing (CTOPP)	6 (3%)	5 (3%)	1 (2%)
Woodcock-Muñoz Language Survey - Revised (WMLS-R)	5 (2%)	0 (0%)	5 (10%)
Bender Visual-Motor Gestalt Test	4 (2%)	4 (3%)	0 (0%)
Wide Range Assessment of Memory and Learning, 2nd Edition (WRAML-2)	4 (2%)	3 (2%)	1 (2%)
Peabody Picture Vocabulary Test, 4th Edition (PVVT-4)	1 (1%)	1 (1%)	0 (0%)
Phonological Awareness Test (PAT)	1 (1%)	1 (1%)	0 (0%)
Test of Memory and Learning, 2nd Edition (TOMAL-2)	1 (1%)	1 (1%)	0 (0%)
Vineland Adaptive Behavior Scales, 2nd Edition (Vineland-2)	1 (1%)	0 (0%)	1 (2%)
Woodcock Language Proficiency Battery (WLPB)	1 (1%)	0 (0%)	1 (2%)
Total number of other assessments identified	10	7	6

Note. Each percentage reflects the proportion of reports in the full sample (N = 212) including that assessment. Data are further reported for the proportion of non-CLD reports (n = 162) and CLD reports (n = 50) that included each assessment. The most recent version of each assessment used is named, but earlier editions were also included for ease of analysis. All tests included in this table were administered by school psychologists as part of the psychoeducational evaluation process. Tests referenced from other reports or evaluations are not included.

The frequency of SLD eligibility subdomains and combinations of these subdomains was computed across the full sample, as well as separately for the CLD and non-CLD subgroups. As depicted in Table 6, for all reports analyzed, the most common eligibility category recommended was Reading Comprehension, which appeared in

approximately half of all cases. Among CLD cases alone, Basic Reading and Reading Comprehension were the most common eligibility categories recommended, with each present in 50% of evaluation reports. Case-by-case eligibility data were also compiled to determine the most common combinations of eligibility categories recommended by clinicians based on evaluation results. Across the full sample, the most frequently observed case was that of a student with sole eligibility in the area of Written Expression (14 cases; 7% of the full sample). This was also true for the non-CLD subgroup (12 cases; 7% of the subsample). However, for CLD students, it was most common for a student to be eligible jointly under the categories of Basic Reading and Reading Comprehension (6 cases; 12% of the subsample).

Table 6

SLD Eligibility Status by Subcategory in Evaluations of CLD and Non-CLD Students

SLD Subcategory	Full Sample (%)	Non-CLD Students (%)	CLD Students (%)
Reading Comprehension	105 (50%)	75 (46%)	30 (60%)
Basic Reading	99 (47%)	69 (43%)	30 (60%)
Written Expression	98 (46%)	77 (48%)	21 (42%)
Mathematics Calculation	80 (38%)	59 (36%)	21 (42%)
Mathematics Problem Solving	73 (34%)	52 (32%)	21 (42%)
Reading Fluency	48 (23%)	31 (19%)	17 (34%)
Oral Expression	15 (7%)	11 (7%)	4 (8%)
Listening Comprehension	15 (7%)	12 (7%)	3 (6%)
Unspecified	10 (5%)	7 (4%)	3 (6%)

Note. Data are reported for the full sample (N = 212), as well as the non-CLD (n = 162) and CLD (n = 50) subgroups.

Descriptive data were also generated on the strategies (including accommodations) and interventions. The full lists of strategies and interventions identified in reports are displayed in Tables 7 and 8, respectively. In total, 44 strategies and 42 interventions were identified in reports. The most frequently cited strategy was use of modified assignments (including chunking/shortening), which was present in nearly half of all reports. Other commonly cited strategies across the full sample included providing extra time on tests or assignments (28%), giving prompts or cues (24%), and providing preferential seating or a change of seat (20%). Roughly 18% of cases included strategies (including accommodations) that were labeled as "interventions", but were determined by the researcher to not actually meet ADE's (2012b) broad definition of interventions. The most common interventions meeting ADE's definition were tutoring (including after school homework clubs and Title I tutoring; 40%), small-group instruction (30%), and SuccessMaker (23%).

Table 7

Instructional Strategies Cited in Evaluations of CLD and Non-CLD Students with SLD

Strategy	Full Sample (%)	Non-CLD Students (%)	CLD Students (%)
Modified assignments (includes chunking/shortening)	40 (47%)	29 (43%)	11 (58%)
Extra time on assignments/tests	24 (28%)	21 (31%)	3 (16%)
Prompts or cues (includes repetition and redirection)	21 (24%)	17 (25%)	4 (21%)
Preferential seating (includes change of seat)	17 (20%)	11 (16%)	6 (32%)
Peer buddy or model	16 (19%)	12 (18%)	4 (21%)
Reinforcement/consequences	10 (12%)	8 (12%)	2 (11%)
Proximity control	9 (11%)	8 (12%)	1 (5%)
Reading tests/materials aloud or modified test format	7 (8%)	5 (8%)	2 (11%)
Parent contact (includes home/school collaboration)	6 (7%)	4 (6%)	2 (11%)

Specialized grouping (includes small class size or program change)	6 (7%)	6 (9%)	0 (0%)
Use of a number line/grid	4 (5%)	3 (5%)	1 (5%)
Use of an agenda/folder or help with organizational strategies	4 (5%)	4 (6%)	0 (0%)
Use of audio books/books on tape	3 (4%)	3 (5%)	0 (0%)
Reviewing homework/tests	3 (4%)	3 (5%)	0 (0%)
Allowing oral answers or dictation of ideas	3 (4%)	2 (3%)	1 (5%)
Allowing the student to retake tests	3 (4%)	1 (2%)	2 (11%)
Consultation with behavior specialist or special education teacher	3 (4%)	2 (3%)	1 (5%)
Check-in system/monitoring sheet	2 (2%)	2 (3%)	0 (0%)
Use of number puzzles/games	2 (2%)	2 (3%)	0 (0%)
Use of manipulatives	2 (2%)	1 (2%)	1 (5%)
Providing written copies of questions/notes	2 (2%)	1 (2%)	0 (0%)
Color coding/highlighting materials	2 (2%)	2 (3%)	0 (0%)
504 plan	2 (2%)	1 (2%)	1 (5%)
Giving work early	1 (1%)	1 (2%)	0 (0%)
Use of online books	1 (1%)	1 (2%)	0 (0%)
Use of lower level textbooks	1 (1%)	1 (2%)	0 (0%)
Use of a spelling dictionary	1 (1%)	1 (2%)	0 (0%)
Allowing breaks	1 (1%)	1 (2%)	0 (0%)
Proofreading writing	1 (1%)	1 (2%)	0 (0%)
Providing choices/behavioral options	1 (1%)	1 (2%)	0 (0%)
Providing the student with his or her own workspace	1 (1%)	1 (2%)	0 (0%)
Asking comprehension questions	1 (1%)	1 (2%)	0 (0%)
Use of an attendance contract	1 (1%)	1 (2%)	0 (0%)
Use of an alphabet line	1 (1%)	1 (2%)	0 (0%)
Use of flash cards	1 (1%)	1 (2%)	0 (0%)
Providing extra practice	1 (1%)	1 (2%)	0 (0%)
Use of a frequency modulation (FM) system	1 (1%)	0 (0%)	1 (5%)
Allowing use of pen rather than pencil	1 (1%)	1 (2%)	0 (0%)
Use of adaptive paper	1 (1%)	1 (2%)	0 (0%)
Allowing the student to call home	1 (1%)	1 (2%)	0 (0%)
Allowing the student to work as a teaching assistant in the library	1 (1%)	1 (2%)	0 (0%)
Providing a second set of books	1 (1%)	1 (2%)	0 (0%)
Office referrals	1 (1%)	0 (0%)	1 (5%)
Retention	1 (1%)	1 (2%)	0 (0%)
Use of thinking maps	1 (1%)	1 (2%)	0 (0%)

Note. Each percentage reflects the proportion of reports citing that particular strategy, within the group of reports that cited strategies, $n = 87$ (excluding reports without strategies). Data are further reported for the proportion of non-CLD reports ($n = 68$) and CLD reports ($n = 19$) that cited interventions. As noted previously, strategies included accommodations.

Table 8

Interventions Cited in Evaluations of CLD and Non-CLD Students with SLD

Intervention	Full Sample (%)	Non-CLD Students (%)	CLD Students (%)
Tutoring (includes homework club and Title I tutoring)	58 (40%)	39 (37%)	19 (49%)
Small-group instruction	43 (30%)	33 (31%)	10 (26%)
SuccessMaker[a]	34 (23%)	27 (26%)	7 (18%)
Modified curriculum/program for a specific academic area (includes tiered pull-out support and resource support)	29 (20%)	23 (22%)	6 (15%)
1:1 instruction (includes 1:1 tutoring)	27 (19%)	18 (17%)	9 (23%)
Support/modeling/assistance/help in an academic area (no specific program)	24 (17%)	18 (17%)	6 (15%)
Reteaching (regardless of group size)	23 (16%)	19 (18%)	4 (10%)
Headsprout	17 (12%)	13 (12%)	4 (10%)
Wilson Reading (includes Fundations)[a]	13 (9%)	12 (11%)	1 (3%)
Guided reading or guided instruction	11 (8%)	10 (9%)	1 (3%)
Working with a particular staff member (e.g., teacher, aide)	11 (8%)	9 (9%)	2 (5%)
English Immersion Studies (EIS) instruction	10 (7%)	0 (0%)	10 (26%)
Positive behavior interventions and supports (PBIS)/behavior intervention plan	9 (6%)	8 (8%)	1 (3%)
Accelerated Reader[a]	8 (6%)	7 (7%)	1 (3%)
Research-based/evidence-based program in a specific academic area (program not specified)	8 (6%)	3 (3%)	5 (13%)
Florida Center for Reading Research (FCRR) methods[a]	5 (3%)	3 (3%)	2 (5%)
Additional instructional time on skills outside the curriculum (includes use of supplemental material and differential learning approaches)	5 (3%)	3 (3%)	2 (5%)
Computer-based intervention (program not specified)	5 (3%)	3 (3%)	2 (5%)
Read180/Read Live[a]	4 (3%)	2 (2%)	2 (5%)
FastMath	4 (3%)	4 (4%)	0 (0%)
Soar to Success[a]	3 (2%)	2 (2%)	1 (3%)
Summer school	3 (2%)	1 (1%)	2 (5%)
Mental health prevention services	3 (2%)	3 (3%)	0 (0%)
System 44	2 (1%)	1 (1%)	1 (3%)
RAZ Kids	2 (1%)	2 (2%)	0 (0%)
Compensatory education	2 (1%)	0 (0%)	2 (5%)
Earobics[a]	2 (1%)	1 (1%)	1 (3%)
6-Minute Solution[a]	1 (1%)	1 (1%)	0 (0%)
Teaching test-taking strategies	1 (1%)	1 (1%)	0 (0%)

Great Leaps[a]	1 (1%)	1 (1%)	0 (0%)
Handwriting Without Tears	1 (1%)	1 (1%)	0 (0%)
Seeds of Change	1 (1%)	1 (1%)	0 (0%)
Touch Math	1 (1%)	1 (1%)	0 (0%)
Words their Way	1 (1%)	1 (1%)	0 (0%)
Science Research Associates (SRA) Math[a]	1 (1%)	1 (1%)	0 (0%)
Harcourt Math	1 (1%)	1 (1%)	0 (0%)
Harcourt Reading	1 (1%)	1 (1%)	0 (0%)
Math Triumphs	1 (1%)	1 (1%)	0 (0%)
Write from the Beginning	1 (1%)	1 (1%)	0 (0%)
Multimodal instruction	1 (1%)	0 (0%)	1 (3%)
Bernie's Typing	1 (1%)	0 (0%)	1 (3%)
Speech services	1 (1%)	0 (0%)	1 (3%)

Note. The broad ADE (2012b) definition of an intervention was used for coding purposes. Each percentage reflects the proportion of reports citing that particular intervention, within the group of reports that cited interventions, $n = 145$ (excluding reports without interventions). Data are further reported for the proportion of non-CLD reports ($n = 106$) and CLD reports ($n = 39$) that cited interventions.
[a]Interventions considered empirically based for the purposes of this study, as defined by WWC (2014) and/or FCCR (2014) criteria.

Finally, descriptive data was generated on the specific types of universal screening and progress monitoring data identified in reports. This data is displayed in Tables 9 and 10, respectively. Within reports containing universal screening data, it was most common to include oral reading fluency scores (words per minute; 74%), followed by math calculation scores (digits correct; 30%), or letter-sound identification scores (22%). The mean number of universal screening data points in reports containing this type of data was 3.71, with a range of 1 to 14. Within reports containing progress monitoring data, it was most common to include oral reading fluency scores (44%), a child's Developmental Reading Assessment (DRA) level (19%), or performance on a consonant-vowel-consonant (CVC) word-reading task (14%).

Table 9

Types of Universal Screening Data Reported in Evaluations of CLD and Non-CLD Students with SLD

Data Type	Full Sample (%)	Non-CLD Students (%)	CLD Students (%)
Oral reading fluency (words per minute)	20 (74%)	16 (76%)	4 (67%)
Math calculation (digits correct)	8 (30%)	6 (30%)	2 (33%)
Letter sound identification	6 (22%)	6 (30%)	0 (0%)
Developmental Reading Assessment (DRA) level	4 (15%)	3 (14%)	1 (17%)
Basic Reading Inventory (BRI)	3 (11%)	3 (14%)	0 (0%)
Sight words	2 (7%)	2 (10%)	0 (0%)
Letter recognition	1 (4%)	1 (5%)	0 (0%)
Mazes	1 (4%)	1 (5%)	0 (0%)

Note. Each percentage reflects the proportion of reports including that type of universal screening data, within the group of reports that included universal screening data, $n = 27$ (excluding reports without universal screening data). Data are further reported for the proportion of non-CLD reports ($n = 21$) and CLD reports ($n = 6$) that cited each type of universal screening data.

Table 10

Types of Progress Monitoring Data Reported in Evaluations of CLD and Non-CLD Students with SLD

Data Type	Full Sample (%)	Non-CLD Students (%)	CLD Students (%)
Oral reading fluency (words per minute)	16 (44%)	10 (37%)	6 (67%)
Developmental Reading Assessment (DRA) level	7 (19%)	6 (22%)	1 (11%)
Consonant-vowel-consonant (CVC) words	5 (14%)	3 (11%)	2 (22%)
Sight words	4 (11%)	4 (15%)	0 (0%)
SuccessMaker data	4 (11%)	3 (11%)	1 (11%)
Basic Reading Inventory (BRI)	3 (8%)	3 (11%)	0 (0%)
Headsprout data	3 (8%)	3 (11%)	0 (0%)
Standardized Test for the Assessment of Reading (STAR) level	3 (8%)	3 (11%)	0 (0%)
Fountas & Pinnell phonological word lists	2 (6%)	1 (4%)	1 (11%)
Letter recognition	2 (6%)	0 (0%)	1 (11%)
Number recognition	2 (6%)	2 (7%)	0 (0%)
Accelerated Reader data	1 (3%)	1 (4%)	0 (0%)
Comprehension questions	1 (3%)	0 (0%)	1 (11%)
Counting	1 (3%)	1 (4%)	0 (0%)
Dynamic Indicators of Basic Early Literacy Skills (DIBELS) vocabulary data	1 (3%)	0 (0%)	1 (11%)
DIBELS comprehension data	1 (3%)	0 (0%)	1 (11%)
Fountas & Pinnell benchmark books	1 (3%)	0 (0%)	1 (11%)
Letter sounds	1 (3%)	1 (4%)	0 (0%)
Lexile scores	1 (3%)	1 (4%)	0 (0%)
Read180 data	1 (3%)	1 (4%)	0 (0%)
Science Research Associates (SRA) math data	1 (3%)	1 (4%)	0 (0%)
Vocabulary words	1 (3%)	0 (0%)	1 (11%)
Wilson Reading data	1 (3%)	1 (4%)	0 (0%)

Note. Each percentage reflects the proportion of reports including that type of progress monitoring data, within the group of reports that included progress monitoring data, $n = 36$ (excluding reports without universal screening data). Data are further reported for the proportion of non-CLD reports ($n = 27$) and CLD reports ($n = 9$) that cited each type of progress monitoring data.

Compliance Data

The present study also generated data on adherence to legal criteria in psychoeducational evaluation reports. A summary of the compliance data generated is displayed in Table 11. Four general areas of compliance were addressed on the coding protocol: the review of existing data, overall disability rule-outs, assessment tools, and the statement of SLD and SLD-specific rule-outs. For ease of comparison, one item from the Language section of the protocol was included with the assessment-oriented items because it dealt with testing.

Table 11

Legal Compliance Standards Addressed in Reports

Legal Criterion	Full Sample (%)	Non-CLD Students (%)	CLD Students (%)
Review of existing data (ARS §300.305)			
Review of existing evaluations	149 (70%)	114 (70%)	35 (70%)
Review of other information provided by parents	209 (99%)	162 (100%)	47 (94%)
Review of current classroom-based performance	197 (93%)	152 (94%)	45 (90%)
Review of local and state-wide assessments	121 (57%)	95 (59%)	26 (52%)
Classroom-based observations by the examiner	158 (75%)	120 (74%)	38 (76%)
Observations from teacher/other service providers	164 (77%)	124 (77%)	40 (80%)
Relevant functional information	210 (99%)	161 (99%)	49 (98%)
Relevant developmental information	206 (97%)	159 (98%)	47 (94%)
Overall disability rule-outs (ARS §300.306-300.307)			
Lack of appropriate instruction in reading	207 (98%)	162 (100%)	45 (90%)
Evidence that the child was provided with appropriate instruction in regular education settings	91 (43%)	70 (43%)	21 (42%)
Data-based documentation of repeated assessments of achievement at reasonable intervals	75 (35%)	58 (36%)	17 (34%)
Lack of appropriate instruction in math	203 (96%)	159 (98%)	44 (88%)
Evidence that the child was provided with appropriate instruction in regular education settings	44 (21%)	34 (21%)	6 (12%)
Data-based documentation of repeated assessments of achievement at reasonable intervals	33 (16%)	25 (15%)	8 (16%)
Limited English proficiency	202 (95%)	152 (94%)	50 (100%)
Description of proficiency level	35 (17%)	0 (0%)	35 (70%)

Assessment tools (ARS §300.304)			
Uses a variety of assessment tools and strategies	209 (99%)	160 (99%)	49 (98%)
Evaluation materials selected/administered so as not to be discriminatory on a racial/cultural basis	127 (60%)	100 (62%)	27 (54%)
Description of racial/cultural background	3 (1%)	0 (0%)	3 (6%)
Rationale for selection/administration of procedures	43 (20%)	31 (19%)	12 (24%)
Evaluation administered in child's native language or other mode of communication and in a form most likely to yield accurate information	202 (95%)	162 (100%)	39 (78%)
Statement of SLD and SLD rule-outs (ARS §300.311)			
Whether the child has SLD	212 (100%)	162 (100%)	50 (100%)
Based on severe discrepancy	209 (97%)	159 (98%)	50 (100%)
Based on RTI	6 (3%)	4 (3%)	2 (4%)
Basis for making the determination	168 (79%)	132 (82%)	36 (72%)
Relevant behaviors noted during the observation	200 (94%)	154 (95%)	46 (92%)
Educationally relevant medical findings	196 (93%)	151 (93%)	45 (90%)
Whether the child does not achieve adequately for his/her age or meet state-approved grade-level standards	210 (99%)	162 (100%)	48 (96%)
Whether the child does not make progress sufficient to meet age or state-approved grade level standards	128 (60%)	95 (59%)	33 (66%)
Pattern of strengths and weaknesses	185 (87%)	148 (91%)	37 (74%)
Rule-out: Vision disability	209 (99%)	160 (99%)	49 (98%)
Discussion of vision factors	203 (96%)	156 (96%)	47 (94%)
Rule-out: Hearing disability	209 (99%)	160 (99%)	49 (98%)
Discussion of hearing factors	203 (99%)	156 (96%)	47 (94%)
Rule-out: Motor disability	89 (42%)	74 (46%)	15 (30%)
Discussion of motor factors	55 (26%)	47 (29%)	8 (16%)
Rule-out: Mental retardation/intellectual disability	211 (99%)	162 (100%)	49 (98%)
Discussion of intellectual factors	210 (99%)	161 (99%)	49 (98%)
Rule-out: Emotional disturbance	168 (79%)	133 (82%)	35 (70%)
Discussion of emotional factors	122 (58%)	98 (61%)	24 (48%)
Rule-out: Cultural factors	186 (88%)	143 (88%)	43 (86%)
Discussion of cultural factors	4 (2%)	1 (1%)	3 (6%)
Rule-out: Environmental/economic disadvantage	189 (89%)	143 (88%)	46 (92%)
Discussion of environmental/economic factors	1 (1%)	0 (0%)	1 (2%)

Note. Data are reported for the full sample (N = 212), as well as the non-CLD (n = 162) and CLD (n = 50) subgroups.

For each report, data on compliance items were totaled to generate an overall proportion of compliance criteria met by the report. Across the full sample, the mean compliance proportion was .86, indicating that, on average, reports addressed 86% of legal criteria. The mean compliance proportion for reports completed on CLD students

was .83, and the mean compliance proportion for reports completed on non-CLD students was .87. As Shapiro-Wilk tests confirmed that the data were not normally distributed, $S\text{-}W = .87$, $df = 212$, $p < .001$, the Mann-Whitney U was used to test for differences between groups. This difference was not statistically significant, $U = 3482.00$, $p = .13$.

Within the review of existing data portion, the most common criterion reports failed to meet was the inclusion of local and statewide assessment data; only 57% of reports included this information. A majority of reports included statements ruling out lack of appropriate instruction in reading and math. However, fewer than half of reports included support for these statements in the form of evidence that the child was provided with appropriate general education instruction (such as information on curriculum) or documentation of repeated achievement measures given at regular intervals (such as classroom benchmark scores or progress monitoring data). As the compliance data outlined in Table 11 indicates, it was more common for such evidence or documentation to be present for reading than for math. Finally, the vast majority of reports (95%) addressed English proficiency as a rule-out and, within the CLD subgroup, about 70% included a description of the child's proficiency level.

Though most reports used a variety of instruments, it was relatively common for reports to fail to state that materials were selected and administered to avoid discrimination on a racial/cultural basis. Only 60% of reports included such a statement. Support for this statement was even rarer, with about 20% of reports including detailed rationale for selection and administration of procedures, and 1% describing the child's racial/cultural background with more than just a single listing of ethnicity.

Within the section on SLD specifically, the most common criterion that reports failed to address was motor disability, with only 42% of all reports including a statement about this, and even fewer including information to support this rule-out. A second criterion that many reports failed to address was the issue of whether the child failed to make progress sufficient to meet age- or grade-level standards. Though all but two reports stated that the child was failing to achieve adequately, only 60% addressed the progress issue. Finally, all reports contained a statement that the child met eligibility criteria. The vast majority (99%) stated that a severe ability/achievement discrepancy was present. Six reports (3%) used the child's failure to respond to evidence-based intervention as justification for the eligibility recommendation; of these six, three also reported an ability-achievement discrepancy and three made the determination solely based on failure to respond.

First Research Question

The first research question addressed how the inclusion of RTI data affects the psychoeducational evaluation process. This question addressed data obtained from full sample of reports included in the study.

It was first anticipated that reports completed within the last two years would be more likely to reference RTI than reports completed within the first two years. This hypothesis was confirmed. Seventy-nine percent of reports from the first two years and 92% of reports from the last two years included statements that prereferral strategies (including accommodations) and/or interventions were attempted prior to evaluation of the student; the relationship between these variables was statistically significant, $\chi^2(1,$

212) = 6.38, p = .01. Moreover, a statistically significant relationship emerged between the listing of specific strategies or interventions and the timing of the report (2009-2010 and 2010-2011 versus 2011-2012 and 2012-2013 school years), $\chi^2(1, 212) = 15.81$, $p <$ .001. Whereas about 65% of reports completed within the first two years included listings of strategies or interventions, 90% of reports within the final two years met this broad criteria.

The second hypothesis was that among reports stating that prereferral strategies and interventions had been attempted, only a few would include concrete information on specific empirically-based interventions, and that many would list accommodations and strategies rather than interventions as they have been defined by the state department of education. It was further expected that reports from the most recent two years would be more likely to reference research-based interventions than reports from the initial two years. This hypothesis was partially confirmed. Using WWC (2014) or FCCR (2014) criteria, 11 of the 42 interventions cited in reports were identified as empirically based. Across the full sample, 84% of reports made the statement that strategies and/or interventions were attempted prior to evaluation. Within this set, 49% included at least one strategy (including accommodations), and 81% included at least one intervention. Thus, the inclusion of interventions was actually more common relative to the inclusion of strategies. However, within that 81% containing interventions, only 37% of reports contained at least one specific empirically-based intervention; this proportion rose to 40% when unnamed empirically-based interventions were included (that is, interventions described as "research-based" or "empirically-based" but not named). When this subset

is considered in relation to the full sample of all reports in the study, less than one-third of all reports (31%) cited the use of specific empirically-based programs. Thus, while it was relatively common for practitioners to report that some sort of instruction had been directed to address areas of concern for the student, it was less common for that instruction to be in the form of a specific empirically-based program. Contrary to expectations, there was no significant difference in the inclusion of empirically-based interventions between the first two years of reports and the last two years, $\chi^2(1, 178) = .01$, $p = .93$.

The third hypothesis was that only a small number of reports referencing prereferral strategies and interventions would include actual RTI data, defined as data from universal screenings and/or progress monitoring. This hypothesis was largely confirmed, as 16% of these reports included universal screening data, and 20% contained progress monitoring data. No significant differences were found in the inclusion of either of these types of data when the first two years were considered against the second two years, $\chi^2(1, 178) = 2.38$, $p = .12$ and $\chi^2(1, 178) = .86$, $p = .35$, respectively. The mean number of universal screening data points in reports containing this type of data was 3.71, with a range of 1 to 14. The mean number of progress monitoring data points in reports containing this type of data was 5.53, with a range of 1 to 19. As Shapiro-Wilk tests confirmed that data on the number of universal screening data points ($S\text{-}W = .81$, $df = 28$, $p < .001$) and the number of progress monitoring data points ($S\text{-}W = .83$, $df = 36$, $p < .001$) violated the normality assumption for traditional t-tests, the Mann-Whitney U test was used. For reports containing universal screening data, there was no significant

difference in the number of data points per report between the first and last two years of reports, $U = 143.50$, $p = .09$. However, for reports containing progress monitoring data, there were significantly more data points present in those written from 2011-2013 than those written from 2009-2011, $U = 240.00$, $p = .01$.

The fourth hypothesis was that reports completed on elementary school students would incorporate more information on the RTI process than reports completed on secondary students. Reports on elementary school students contained a greater mean number of universal screening data points ($U = 3536.50$, $p = .004$), progress monitoring data points ($U = 3433.00$, $p = .003$), and interventions ($U = 3169.00$, $p = .002$) than reports completed on secondary students. Universal screening and progress monitoring data were extremely uncommon in reports completed on secondary students; within this subgroup, there was just one report containing universal screening and progress monitoring data, and a second containing progress monitoring data. There was insufficient data to analyze the specific length (e.g., months), frequency (e.g., daily), and timing (e.g., minutes per day) of interventions, due to the vast majority of reports using nonspecific terminology to specify how long an intervention had taken place (e.g., "since first grade"). In lieu of this, the relative proportions of reports containing data on the length, frequency, and timing per intervention session were compared for elementary and secondary students. Due to the small number of data points in some cells, Fisher's exact tests were used over the chi-square test to analyze data on the inclusion of the length and timing of intervention; reports on elementary school students were significantly more likely to include information on the length of interventions, $p = .004$, and the timing of

interventions, $p = .03$. However, reports on elementary school children were not significantly more likely to report the frequency of interventions, $\chi^2(1, 209) = 2.82$, $p = .09$.

The fifth hypothesis tied to this research question was that significantly more reports would contain RTI data in the area of reading than in any other academic domain. Although all school psychologists in the district had access to curriculum-based measures in writing, no reports in the sample included universal screening or progress monitoring data in writing. One report referenced use of a writing CBM, but treated that measure as an assessment administered alongside traditional academic tests, rather than as a universal screening or progress monitoring tool. A Fisher's exact test indicated the relationship between the inclusion of universal screening data in reading and the inclusion of universal screening in math across all reports was statistically significant, $p < .001$. Of the 27 reports containing universal screening data, 26 (96%) contained data in reading and 8 (30%) contained data in math. A Fisher's exact test also demonstrated a statistically significant relationship between the inclusion of progress monitoring data for reading and progress monitoring data for math across all reports, $p = .001$. Of the 36 reports containing progress monitoring data, all contained progress monitoring data in reading, and 4 (11%) included progress monitoring data in math.

The sixth and final hypothesis was that reports referencing the RTI process (including prereferral strategies/interventions and data generated through these) would be likely to use fewer traditional assessments than reports without this information. This hypothesis was not supported by the data. Shapiro-Wilk tests confirmed that data on the

overall number of assessments used per report were not normally distributed, *S-W* = .93, *df* = 212, *p* < .001. Data on the number of cognitive (*S-W* = .54, *df* = 212, *p* < .001), academic (*S-W* = .64, *df* = 212, *p* < .001), social/emotional/behavioral (*S-W* = .82, *df* = 212, *p* < .001), and other (*S-W* = .57, *df* = 212, *p* < .001) assessments also violated normality assumptions. Thus, the Mann-Whitney U was used. No relationship was found between the total number of traditional assessments per report and the presence or absence of prereferral strategies/interventions, *U* = 4948.50, *p* = .10. There were also no relationships present when cognitive (*U* = 4829.50, *p* = .05), academic (*U* = 4007.00, *p* = .31), social/emotional/behavioral (*U* = 4853.00, *p* = .15), and other (*U* = 4546.00, *p* = .45) tests were considered separately. Likewise, no relationship was found between the total number of traditional tests per report and the inclusion of universal screening data (*U* = 2913.50, *p* = .25) or the inclusion of progress monitoring data (*U* = 3522.50, *p* = .28). Comparisons of academic (*U* = 2312.50, *p* = .27; *U* = 3342.00, *p* = .51), social/emotional/behavioral (*U* = 2675.00, *p* = .73; *U* = 3149.50, *p* = .95), and other (*U* = 2776.50.00, *p* = .39; *U* = 3493.50, *p* = .21) tests also yielded no significant results. However, there were significant relationships between the number of cognitive tests given and the number of universal screening points (*U* = 3505.00, *p* < .001), as well as between the number of cognitive tests given and the number of progress monitoring points (*U* = 3645.00, *p* = .03). Reports containing these types of data tended to include a greater number of cognitive measures than reports without these types of data.

Second Research Question

The second research question addressed how the psychoeducational evaluation process differs for CLD students, particularly with regard to the inclusion of RTI data.

The first hypothesis posited that there would be greater variability in the measures in evaluations completed for CLD students than evaluations completed for non-CLD students. This was assessed in two ways – by comparing the mean number of assessments (overall and in each domain) per report, as well as the total number and range of assessments used across each subsample. Tables 2-5 display the full array of assessments identified in reports, while per-report descriptive data (including the mean, standard deviation, mode, median, and range of tests used) for CLD and non-CLD students is displayed in Table 12. Overall, this hypothesis was not supported. As noted previously, data on the number of assessments (total and per domain) used per report were not normally distributed. Consequently, the Mann-Whitney U was used. There was no significant difference in the total number of tests used per report for CLD and non-CLD students, $U = 3827.50$, $p = .55$. There were also no differences in the number of cognitive ($U = 4135.50$, $p = .74$), academic ($U = 4331.50$, $p = .35$), social/emotional/behavioral ($U = 3553.00$, $p = .16$), or other ($U = 3897.00$, $p = .61$) assessments used per report for CLD and non-CLD students. Though there were some minor differences in the ranges for different types of tests used, overall, the standard deviations associated with each type of test were very similar for both CLD and non-CLD students. Further, the total numbers of unique cognitive, academic, and other assessments identified were comparable across the CLD and non-CLD groups. Although

the non-CLD evaluations as a group contained more than twice as many unique social/emotional/behavioral measures as the CLD group, this would be expected given the difference in subsample sizes.

Table 12

Descriptive Statistics for Assessment Types Used in Evaluations of CLD and Non-CLD Students with SLD

Assessment Type	Non-CLD Students				CLD Students			
	M (SD)	Median	Mode	Range	*M (SD)*	Median	Mode	Range
Cognitive assessments	1.14 (.40)	1	1	0-2	1.18 (.48)	1	1	0-3
Academic assessments	1.32 (.64)	1	1	0-4	1.38 (.64)	1	1	0-3
Social/ emotional/ behavioral assessments	.97 (.99)	1	0	0-5	.76 (.92)	1	0	0-3
Other assessments	.31 (.54)	0	0	0-3	.32 (.68)	0	0	0-3
All assessments	3.75 (1.48)	4	3	0-8	3.64 (1.64)	3	2	0-8

Note. Data presented indicate assessments used per report. As shown in Table 5, other assessments included a variety of measures not captured by other categories, including memory, adaptive, visual-motor, and language dominance tests,

The second hypothesis was that cognitive assessment of CLD students would be more likely to rely upon nonverbal measures of ability than assessment of non-CLD students. Evaluations of CLD students were significantly more likely to include distinctly nonverbal measures of ability, $\chi^2(1, 212) = 17.93$, $p < .001$, than evaluations of non-CLD students. Moreover, evaluations of CLD students were more likely to include

only nonverbal measures in lieu of full cognitive assessments, $\chi^2(1, 212) = 10.33$, $p = .001$.

The third hypothesis contained two parts. First, it postulated that, within the CLD subgroup, nonverbal assessments would be the most commonly used measures, followed by traditional cognitive measures given in English, and finally by cognitive measures given in Spanish. Second, it postulated, that within this subgroup, academic assessments in English would be more commonly used than academic assessments in Spanish. This hypothesis was only partially confirmed. As depicted in Table 2, the two most commonly-used cognitive measures for CLD students – the WISC-IV and full KABC-2 – were also the most common cognitive assessments for non-CLD students. The KABC-2 Nonverbal Index was the third most commonly used measure for both subgroups. Cognitive measures administered in Spanish were exceedingly rare, with only two reports in the entire sample including them. As shown in Table 3, achievement tests in English were the most commonly used academic measures. The Batería III Woodcock-Muñoz Achievement Battery was the only Spanish achievement measure identified, and it was used in only four evaluations (8% of the CLD subgroup).

The fourth hypothesis posited that evaluations of CLD students would be more likely than evaluations of non-CLD students to go beyond formal IQ and achievement testing to include additional pieces of evidence to confirm that the child's failure to learn is unrelated to his or her language status. It was hypothesized that these would include information on prereferral strategies, interventions, and RTI data, but also data on progress within the general curriculum, such as benchmark scores. This hypothesis was

not confirmed. Reports completed on CLD students were no more likely to state that prereferral strategies and interventions had been attempted prior to evaluation, $\chi^2(1, 212) = 1.77$, $p = .18$, nor were they more likely to name specific strategies or interventions that were tried, $\chi^2(1, 212) = 1.20$, $p = .27$. Similarly, there were no significant differences in the number of interventions attempted ($U = 4554.00$, $p = .17$), the amount of universal screening data ($U = 3976.00$, $p = .74$) or the amount of progress monitoring data ($U = 4125.50$, $p = .76$) in reports completed on CLD and non-CLD students. Only two reports formally used failure to respond to intervention as part of their eligibility determination, and in both of these cases, an ability-achievement discrepancy model was also identified. Finally, reports on CLD students were no more likely to contain local and statewide assessment data, including scores on benchmark tests, $\chi^2(1, 212) = .69$, $p = .41$.

The final hypothesis within this cluster predicted that the methods by which school psychologists ruled out linguistic proficiency as the main factor in learning challenges would be extremely variable. It was expected that parent input, educational history, state proficiency test data, and scores from individually-administered proficiency tests would be included. Proportional data on broad factors that relate to exclusionary criteria, but were legally-required components of all reports (e.g., parent input, classroom observations) can be found in Table 11. Methods by which school psychologists addressed linguistic proficiency issues specifically are displayed in Table 13.

Table 13

Factors Used to Address Linguistic Proficiency in Evaluations of CLD Students with SLD

Quality	Reports Including (%)
Home language specified	41 (84%)
Spanish	24 (59%)
English	8 (20%)
English/Spanish	6 (15%)
Urdu	1 (2%)
Romanian	1 (2%)
English/Navajo	1 (2%)
English/French	1 (2%)
Primary language of student specified	41 (84%)
English	21 (51%)
Spanish	11 (27%)
English/Spanish	10 (24%)
Educational history included	40 (80%)
Individually-administered proficiency test referenced in, but not administered as part of, this evaluation (e.g., from speech evaluation, from previous evaluation)	19 (38%)
Woodcock-Muñoz Language Survey – Revised (WMLS-R)	11 (58%)
Bilingual speech/language evaluation (separate)	5 (26%)
Proficiency test (nonspecific)	2 (11%)
Language screening (nonspecific)	1 (5%)
Stanford English Language Proficiency Test (SELP)	1 (5%)
State proficiency test (AZELLA) score stated	16 (33%)
Individually-administered proficiency test administered as part of the evaluation	12 (24%)
Woodcock-Muñoz Language Survey – Revised (WMLS-R)	7 (58%)
Bilingual speech/language assessment included in report	1 (8%)
Woodcock Language Proficiency Battery (WLPB)	1 (8%)
Student's first language specified	8 (16%)
English	2 (25%)
Spanish	4 (50%)
English/French	1 (13%)
Luganda	1 (13%)
Years living in U.S. reported	7 (14%)
Uses different model of test interpretation based on language history (e.g., use of Luria model to interpret KABC-2 results)	7 (14%)
Additional languages spoken in the home specified	5 (10%)
Non-English achievement test administered	4 (8%)
Uses child's failure to respond to intervention to rule out linguistic proficiency as primary factor in learning difficulties	2 (4%)
States that child appeared to understand directions equally in both Spanish and English during assessment	1 (2%)
States that child speaks only English at school	1 (2%)

Note. Percentages within categories indicate the proportion of reports within that category meeting the criteria (e.g., the percentage of reports citing the child's home language in which that home language was Spanish).

All of the expected areas were addressed in at least some reports, but to varying degrees. Most reports included parent input (94%), stated the home language of the student (84%), stated the primary language of the student (84%), and discussed the educational history of the student (80%). It was far less common to specify the student's first language, with only 8% of reports doing so.

Twelve of the fifty reports completed on CLD students (24%) included data from a language proficiency test that was administered individually to the child as part of the evaluation process. The most common of these measures was the Woodcock-Muñoz Language Survey - Revised (WMLS-R; Woodcock & Muñoz-Sandoval, 2011). An additional 19 reports (38%) were given separately from the psychoeducational evaluation (prior to it or as part of a speech/language evaluation), but were cited in the psychoeducational evaluation report. Additionally, one-third of reports included information on state proficiency test (AZELLA) scores.

It was further expected that these trends in assessment of CLD students may relate to the school psychologists' bilingual status. This hypothesis was confirmed, but in an unexpected direction. A Fisher's exact test indicated a statistically significant relationship between the use of nonverbal cognitive tests and the bilingual status of the school psychologist, $p < .001$, with bilingual psychologists found to actually be more likely to use nonverbal cognitive measures with CLD students than monolingual psychologists. Moreover, a Fisher's exact test also indicated that bilingual school psychologists were more likely than their monolingual counterparts to rely on a nonverbal test as the only cognitive measure for a CLD student, $p = .004$. The two

reports containing data from cognitive measures completed in Spanish were also completed by bilingual school psychologists, as were the four reports containing data from academic measures completed in Spanish. There was also a statistically significant relationship between the administration of individual proficiency tests as part of the evaluation and the bilingual status of the psychologist, with bilingual practitioners again more likely to include these assessments as part of their evaluations, $\chi^2(1, 50) = 7.74$, $p = .005$. However, a Fisher's exact test indicated no relationship between the practitioner's bilingual status and the likelihood of including proficiency test data from another evaluation, $p = 1.00$. No school psychologists, bilingual or monolingual, reported having used an interpreter during assessment.

CHAPTER 4

DISCUSSION

Research Summary

Since its endorsement in IDEIA (2004), RTI has become a focus of extensive research within the field of school psychology. Though a wealth of literature exists on psychometric issues and best practices within the context of RTI models of service delivery (Fuchs & Fuchs, 2006; Ikeda et al., 2008; Shinn, 2002), less research has been done on the practical issues surrounding how data generated through the RTI process may be incorporated into the psychoeducational evaluation process for students suspected of having learning disabilities. Moreover, although RTI is often touted as a method of helping effectively identify struggling CLD students and ruling out the role of linguistic proficiency in their difficulties (Kamps et al., 2007), little research exists on whether or not practitioners are actually using RTI data in that manner, and if so, how they are doing it.

The present study addressed two main research questions: how RTI influences the psychoeducational evaluation process, and how this process may differ for CLD and non-CLD students. The study's data source was psychoeducational evaluation reports completed on 212 students (162 non-CLD and 50 CLD) who were found eligible for special education services under the SLD category from 2009-2013. A detailed coding scheme was developed. Prior to this research, only one study had addressed the content of evaluation reports completed on CLD students (Wilkinson et al., 2006); that study had a limited sample and did not allow for comparison of practices with CLD and non-CLD

students. The present study aimed to fill that gap in the literature by exploring this topic with a larger and broader sample, while at the same time generating extensive descriptive information on general assessment practices and adherence to legal criteria in reports.

Integration of RTI Data

Conclusions. Results from analysis completed on the full sample of reports suggested that practitioners are increasingly integrating information on prereferral strategies/interventions into psychoeducational evaluations, but not always with great specificity. Reports completed within the most recent two years were more likely to state that strategies and/or interventions were tried prior to evaluation, as well as more likely to cite examples of these approaches. Yet despite the fact that more than 80% of reports in the present study mentioned having provided some sort of "directing of instruction in the area(s) of concern" (ADE, 2012b, p. 4) – the definition of an intervention put forth by the state department of education – less than a third cited specific methods that could be verified as empirically-based. This trend was no different between the present study's subsets of older and more recent reports. Thus, within the district of focus for the study, while it may be becoming more common for practitioners to cite prereferral efforts, there has not necessarily been a move toward the inclusion of more empirically-based methods.

Similarly, results from the present study indicated that a minority of reports included universal screening or progress monitoring data, with no change in the proportions of reports including these types of data over time. If a practitioner did include universal screening data in the report, there were about 3-4 data points present on average; for progress monitoring data, that figure was greater at about 5-6. Interestingly,

newer reports with progress monitoring data had, on average, a greater quantity of that data than older reports with progress monitoring data. It is possible that this difference may be reflective of the greater amount of data available to practitioners as time progresses and schools accumulate more and more data on student progress in interventions. At the same time, though, this difference was not present in universal screening data. This may be due to the fact that, within an RTI framework, universal screenings are typically administered for a finite amount of times per year (e.g., fall, winter, and spring; Fuchs & Fuchs, 2006; Jenkins, Hudson, & Johnson, 2007), whereas there is greater flexibility and room for growth with regard to how often students are progress monitored.

Information on the prevalence of RTI data and interventions in reports also suggests that, consistent with the extant literature on RTI, RTI implementation is far more common at the elementary level than at the secondary level. Reports completed on elementary school children contained more universal screening data, progress monitoring data, and interventions than reports completed on older students. They were also more likely to include information on how long interventions were implemented and the duration of each session (e.g., 30 minutes/day). Though the literature on RTI may lead one to expect that certain types of data may be more common in middle and high school than in elementary school (Graney et al., 2010; Tichá et al., 2007; Lembke et al., 2012), in general, the inclusion of any type of RTI data for secondary students was extremely rare within this study's sample – only two reports (4%) included it, and those reports did not include the maze probes and advanced math probes cited in the literature as common

measures for secondary school use. At the elementary level, oral reading fluency scores emerged as the most common type of universal screening and progress monitoring data present in reports by far. This is consistent with current research suggesting that this type of CBM is the most commonly used CBM for elementary school students (Tindal, 2013).

In general, reading data was much more commonly found in reports than math data, a finding that is consistent with the literature on RTI practice (Jenkins et al., 2013). Also consistent with the literature was the relative paucity of information on writing-specific interventions and writing-focused data (Berninger et al., 2008; McMaster & Espin, 2007). Excluding typing and handwriting-only programs, only one writing-specific intervention – Write from the Beginning – was identified in the reports analyzed for the present study. This may be a slight underrepresentation due to the fact that some reading intervention programs also include writing and spelling components (Joseph, 2008). However, coupled with the fact that no writing CBMs were cited as progress monitoring tools across the full sample of reports, it appears that writing is not a target of extensive, structured intervention within the district of focus. At the same time, the most common case in the sample was that of a student with sole eligibility in the area of Written Expression (that is, students who were only eligible in Written Expression and no other areas). Though disabilities in Reading Comprehension were more common overall (in combination with other areas), the preponderance of students with writing-only disabilities, coupled with the lack of focus on writing interventions, prompts questions about how these trends may be related.

Finally, findings from the present study indicated that reports on both CLD and non-CLD students that referenced RTI were not likely to include fewer traditional assessments and, in fact, were likely to include a greater number of cognitive measures than reports with less emphasis on RTI. One reason for this trend may be tied to research suggesting that, when an RTI framework is in place, schools end up evaluating fewer students because more students are having their learning difficulties remediated without having to undergo psychoeducational testing and placement in special education (VanDerHeyden et al., 2007; Sullivan & Long, 2010; Torgesen, 2009). When they are doing fewer evaluations per year, practitioners may be able to spend more time conducting extensive psychoeducational assessments to help determine precisely why children are not learning, and those assessments may include more cognitive measures that help identify specific strengths and weaknesses in reasoning ability.

Implications. Overall, these data provide an illustration of the realities of RTI implementation through the lens of psychoeducational reports, as well as of how those realities do not always coincide with what is recommended in the literature. Within this study's sample, there was significant variability in the prereferral strategies and interventions cited. This variability suggests the absence of a standard protocol approach to RTI within the district of focus. Rather than using the same core set of assessment and intervention materials for all children, school teams may be tailoring strategies specifically to the needs of individual students. Alternatively, the realities of budget constraints and autonomy among schools may have promoted adoption of different tools and strategies by individual campuses within the district. Regardless, although

practitioners tend to prefer the more individualized problem solving approach, the research base for the standard protocol model is superior (Fuchs et al., 2003; Shores & Chester, 2009).

Additional discrepancies between the reality of RTI implementation and recommendations on best practices were revealed through analysis of the specific strategies, interventions, and assessment tools used. Notably, many reports included strategies/accommodations that were referred to as interventions, but did not involve the targeting of instruction that characterizes true interventions. In practice, the lines distinguishing between instructional strategies, accommodations, and true interventions are often blurred (Brown-Chidsey, Bronaugh, & McGraw, 2009; McCook, 2006), and the data from this study support that. Moreover, only a quarter of the true interventions identified in reports could be verified as empirically-based. This is highly problematic because it undermines fundamental assumptions of RTI models (Fuchs & Fuchs, 2005; Yell, Shriner, & Katsiyannis, 2006), as well as the emphasis in federal law on employing instructional strategies that are supported by research (U.S. Department of Education, 2006). Finally, despite the prevalence of literature citing can't do/won't do assessments and decision-making rules for nonresponsiveness as part of best practices in RTI (VanDerHeyden & Witt, 2008; Ardoin et al., 2013), no reports referenced using these. Overall, scholars have speculated that the research-practice gap that plagues special education also extends to the implementation of RTI frameworks (Kozleski & Huber, 2010), and these data provide tentative support for that.

On a practical note, the data generated by the present study may have direct utility for the school district that provided the data. Descriptive information on the assessments being used by practitioners can help inform administrators' purchasing decisions. From a legal standpoint, the information on compliance with special education law highlights areas of strength and weakness in the district's psychoeducational reports. This data may be used by administrators in the development of standard operating procedures, including the revision of existing report templates. Yet perhaps most importantly of all, the data may serve to generate critical discussion among school psychologists about trends within the district and ways that practitioners can all work to better their practice. Issues raised by this study, such as the preponderance of writing-only disabilities and apparent lack of writing interventions, merit further exploration by the district in an effort to better understand the connection between RTI practices and student outcomes.

The coding instrument created may also have practical utility beyond the district of focus. The Office of Special Education Programs (OSEP) is responsible for monitoring states' compliance with IDEIA (2004), and in turn, state departments of education monitor school districts to ensure compliance. Use of a self-monitoring tool such as the one developed for the present study may assist any district with meeting compliance standards, as well as promote ethical and legally compliant evaluation practices. Though the instrument is based on Arizona law, its close correspondence with IDEIA standards would allow it to be easily adjusted to suit standards in any state.

Limitations and future research. A key limitation of the present study stems from the nature of its data source. Psychoeducational evaluation reports present a

valuable vehicle for studying assessment practices because they are direct products of the evaluation process and are required to contain specific components outlined in IDEIA (2004). Differences in the strategies, interventions, and data presented in reports may be indicative of actual differences in the prereferral intervention process for a particular student, but they may also be indicative of the psychologist's report writing style and inclination to include these pieces of information in reports. The specific reason for a child's referral for evaluation (e.g., difficulties in reading and/or difficulties in math) may have also played a role in a psychologist's decision to include data on a child's progress in a particular area; this was not addressed in the study's coding scheme. Additionally, the definition used for coding interventions (ADE, 2012b) was quite vague in nature, despite its basis in state law. This allowed for a wide variety of practices, including tutoring and support or assistance in a particular academic area, to be coded as interventions. In the future, follow-up survey or interview-based research with practitioners may help clarify how RTI is practiced and, more specifically, how it affects a practitioner's approach to evaluating a student for SLD. Additionally, the coding scheme used in the present study may be refined to include items on the presence or absence of specific reasons for a child's referral for evaluation, as well as a narrower definition of an intervention.

A related limitation lies in the fact that reports were retrieved for the study using school psychologists' logs. A significant number of reports that were initially identified on the logs were not included in the final study. Some reports were retrieved from the district's archives, but did not actually turn out to meet study criteria; this was likely due

to the fact that psychologists submitted logs prior to the end of the school year, and therefore were not certain what the ultimate eligibility status of the student would be at the time. However, other reports simply could not be found in the archives, and thus may have never been completed. It is possible that there could have been systematic differences between these reports and the reports included for study that altered the findings. Future studies on this topic would benefit from use of alternative sampling means, such as through district-wide computer databases that are less prone to error; unfortunately these were not made available for the present study.

Additionally, the study's data type and sample size limited the statistical analyses that could be performed. Much of the data necessitated use of nonparametric tests, which tend to have less power in general than their parametric counterparts (Siegel, 1956). Additionally, in several instances, the number of cases in one or more cells was lower than the advisable value needed for chi-square tests; thus, Fisher's exact tests were used. As these tests are more conservative, they are also somewhat less powerful than chi-square tests, so it is possible that some true between-group differences may have gone undetected (McDonald, 2009).

Finally, a major limitation of the study lies in the generalizability of its findings. All reports were drawn from a single school district in the southwestern U. S., which severely limits the ability to apply these conclusions to other districts and states. Future research studies utilizing the same methodology and coding protocol with reports completed in other districts and regions would add confidence to the findings.

Evaluation Practices with CLD Students

Conclusions. Contrary to expectations, the present study did not yield support for the idea that evaluations of CLD students are characterized by greater variability of assessments relative to evaluations of non-CLD students. This contradicts the findings of Ochoa, Powell et al. (1996), who surveyed school psychologists about their practices with bilingual and limited English proficient (LEP) students and concluded that evaluators used a greater variety of assessments with CLD students. Some differences between the present study and the work of Ochoa, Powell et al. are notable and preclude more in-depth comparisons. First, Ochoa and colleagues' sample identified many more unique cognitive, academic, and adaptive measures overall than were identified in the subsample of CLD evaluations in the present study. This is likely due to that study having a larger sample of data from psychologists across the United States, rather than data from a single district as in the present study. Additionally, the work of Ochoa and colleagues was based on a survey of school psychologists and did not include data on social/emotional/behavioral measures. In contrast, the present study considered data on the type and quantity of assessments that were administered, as documented through evaluation reports. Most important of all, Ochoa and colleagues only considered the overall number of assessments, and did not address the mean number of measures per report. With these methodological differences, it is impossible to ascertain whether or not the lack of differences in assessment variability in the present study is indicative of a true change in practice over the past two decades. All that can be determined is that data from the present study suggest that psychologists in the district of focus have not

employed a larger, more varied set of tools to evaluate CLD students versus non-CLD students.

Data from the present study on the specific tests used prompt comparison with the extant literature on common practice. Evaluations completed on CLD students were significantly more likely to include nonverbal measures of ability than reports completed on non-CLD children, a finding that is consistent with the findings of Ochoa, Powell et al. (1996) and Wilkinson et al. (2006). With the present study's focus on case-by-case use of tests, we were able to expand upon that finding to also determine that practitioners were more likely to rely solely upon nonverbal tests – in the absence of other cognitive measures – in evaluations with CLD students. Even so, the top cognitive test used with CLD students in both the Ochoa, Powell et al. study and the present study was the Wechsler (albeit, different editions). In the present study, the KABC-2 full battery and KABC-2 Nonverbal Index emerged as the second and third most common cognitive tests. Though these were also popular in the Ochoa, Powell et al. study, the second and third most common measures identified in that research were nonverbal – the Leiter International Performance Scale and the Draw-A-Person test; in the Wilkinson et al. study, the top three cognitive measures had all been nonverbal. Administration of cognitive measures in Spanish was rarer in the present study than in both the Ochoa, Powell et al. and Wilkinson et al. studies, although this finding is tempered by the fact that the proportion of bilingual-to-monolingual psychologists in these older works was considerably greater relative to the proportion for the current study.

Overall, these data provide tentative support for the idea that trends in the selection of cognitive assessments with CLD students have not changed dramatically over time. Yet the same may not be true of trends in academic assessments. The most commonly used assessment in the Ochoa, Powell et al. (1996) and Wilkinson et al. (2006) studies of practices with CLD students was, by far, the Batería. Use of this measure was rare in the present study, with psychologists clearly favoring English assessments such as the Woodcock-Johnson and WIAT for use with CLD students. As it makes sense to assess children in the language in which they have primarily been taught (Olvera & Gomez Cerrillo, 2011), it is likely that the difference in practice was due to differences in the characteristics of the students with whom each group of psychologists worked. Wilkinson et al. collected data from a state that allows bilingual instruction (Texas), and Ochoa, Powell et al. surveyed a national sample of school psychologists. As the present study's sample was drawn from a state requiring English-only teaching, there were likely fewer children who had been educated in Spanish. Additionally, the fact that the proportion of reports in the present study on CLD students authored by bilingual practitioners was relatively low (16%) is likely a major factor in this trend. At the same time, monolingual practitioners presumably have the ability to seek consultation and assistance from bilingual colleagues if they believe assessment in Spanish would be useful for the student. Thus, the relative rarity of academic assessments in Spanish in the present study, while not necessarily a piece of information that is comparable to extant research, provides useful information on common practice within the district of focus.

A final area of discussion lies in the specific methods used by practitioners to rule out linguistic proficiency status as the chief factor in learning difficulties for CLD students. Contrary to expectations, evaluations of CLD students were no more likely than evaluations of non-CLD students to reference prereferral interventions, nor did they include more information on interventions, data on the student's response to targeted interventions, or data on the student's progress within the general curriculum. Failure to respond to intervention was specifically cited by practitioners as part of the rationale for the eligibility recommendation in only two reports completed on CLD students, and four reports completed on non-CLD students. In all CLD cases, including the two that formally cited RTI failure to respond, practitioners cited the existence of an ability-achievement discrepancy. These findings are consistent with those of Wilkinson et al. (2006) study, which had also been completed in a state (Texas) in which alternatives to the discrepancy method are permitted. In many cases reviewed by the panel in that study, experts expressed concern about the lack of documentation of prereferral interventions, which they felt was significant enough to prompt questions about whether or not the students were truly learning disabled (Wilkinson et al., 2006).

This prompts the question of how psychologists are ruling out linguistic proficiency, if it is not through the inclusion of RTI documentation or data on progress within the general curriculum. Consistent with earlier survey-based research on factors used by psychologists to rule out linguistic proficiency (Ochoa et al., 1997), general parent input regarding the child and family history emerged as the most commonly included piece of information. Proficiency test data was common in the present study,

but it came from a variety of sources. Individually-administered proficiency tests were administered as part of 24% of the CLD evaluations studied; at the same time, in another 38% of CLD reports, psychologists included individually-administered proficiency test data from other sources (speech/language evaluations and prior testing). In total, more than 60% of evaluations included this type of data. Previous research had suggested that this type of data was much rarer; for example, in Ochoa et al.'s (1997) survey of practitioners, only 11% reported using any kind of proficiency test data in their reports, including individually-administered, district-wide, and state-wide measures. Similarly, the home and primary languages of the student were also present in the majority of reports in the current study. Only a small minority of the practitioners surveyed by Ochoa et al. had reported using these factors in addressing exclusionary criteria. Again, though, methodological differences are important to consider. Ochoa et al. asked practitioners what factors they used to address exclusionary criteria; thus, it is possible that most practitioners may have included these pieces of information in their reports, but not reported them as factors to consider in addressing proficiency issues.

Finally, the current study provides limited insight into the relationship between bilingual status of psychologists and trends in evaluation content. Ortiz (2004) has argued that bilingual psychologists are the best practitioners to assess CLD students, because they are able to conduct truly bilingual evaluations. In the present study, bilingual practitioners were, indeed, more likely than their monolingual counterparts to include language proficiency tests as part of their evaluations. Yet somewhat unexpectedly, bilingual practitioners were also more likely to use nonverbal cognitive

tests in their evaluations of CLD students than monolingual practitioners, both solely and in addition to other cognitive tests. Though interesting, this finding must be interpreted with caution due to the fact that the present study sampled reports completed by a single district's team of 21 monolingual and four bilingual psychologists. With such small numbers, findings may be attributable to personal preference among these practitioners. Also interesting was the fact that no psychologists reported having used an interpreter during assessment, a finding that contradicts the work of Ochoa, González et al. (1996) on the frequent use of interpreters. It is possible that this is indicative of a shift in practice within the field, but it is also likely an area of practice that is highly variable from district to district. In general, students whose lack of English proficiency is significant enough to merit use of an interpreter may not be referred for formal assessment due to the assumption that language proficiency cannot be ruled out as the primary cause of their challenges.

Implications. Though it is beyond the scope of this study to make recommendations on assessment practices, it is worth drawing some comparisons between the findings and the current literature on best practices. In their study, Ochoa et al. (1997) made the statement that their research was "especially disheartening because it shows that the majority of school psychologists fail to recognize the significance of language in the educational status of LEP and bilingual students" (p. 165). The findings of the present study provide a more optimistic outlook on assessment practices with CLD students. By and large, evaluations tended to incorporate multiple sources of information, including parent input, teacher input, information on classroom

performance, and at least one classroom observation. In a majority of the evaluations reviewed, psychologists not only reported the home and primary languages of the student, but also sought additional means to understand the child's linguistic proficiency, often through reviewing or administering norm-referenced proficiency measures. In no cases did evaluators report having had interpreters translate assessment materials. These have all been noted as key components of best practice with students from diverse cultural and linguistic backgrounds (Blatchley & Lau, 2010; Olvera & Gomez-Cerrillo, 2011).

Even so, the reliance on nonverbal measures in many reports stands in contrast to what has been recommended as best practice. Nonverbal measures may be less culturally and linguistically loaded than assessments with verbal components, but students' performance on them is still impacted by their language skills, as well as their exposure to the mainstream culture in which the tests were developed (Ortiz, 2004). Nonverbal tests also fail to provide a full picture of the breadth of a child's reasoning skills (Blatchley & Lau, 2010). Related to this assertion is the recent finding that nonverbal assessments are actually weaker predictors of academic achievement in students from diverse linguistic backgrounds than measures of overall reasoning ability (Lakin, 2012). Though a multitude of flaws have been found in the ability-achievement discrepancy model, it remains the dominant model of SLD identification, and was the primary model found in the present study's sample of evaluation reports. A fundamental assumption of that model is that cognitive ability should predict academic achievement; when a child's achievement is unexpectedly low, that child is thought to have a learning disability. Yet if practitioners are continuing to rely on nonverbal measures of ability with CLD

students, and these measures are not the best predictors of academic achievement, this adds an additional threat to the validity of the model.

To date, there exists no single "gold standard" model for assessment of CLD learners. Olvera and Gomez-Cerrillo (2011) have proposed a cross-battery approach that incorporates a thorough file review, curriculum-based assessment, cognitive measures that coincide with domains of intelligence outlined in Cattell-Horn-Carroll (CHC) theory, linguistic proficiency assessment, and academic testing in the language in which the child has been primarily instructed. Others have stressed the need to document prereferral interventions and use CBMs to monitor the gains of CLD students and develop appropriate expectations for their progress (Blatchley & Lau, 2010; Ortiz, 2004; Wilkinson et al., 2006). Yet in the present study, prereferral documentation and progress monitoring data were no more likely to appear in evaluations of CLD students than those of non-CLD students. Most recently, Flanagan et al. (2013) published a Culture-Language Interpretive Matrix (C-LIM) that practitioners may use to consider the cultural and linguistic loading associated with various assessments and how that should affect their interpretation; however, the validity of the model has been questioned (Styck, Watkins, & Vanderwood, 2013). Overall, in the absence of one dominant protocol for CLD assessment, a multifaceted evaluation that considers all potential factors in a student's learning difficulties and how the validity of assessments may be compromised by those factors appears ideal.

Limitations and future research. Conclusions drawn from the present study on the nature of CLD assessment are limited by the same issues of data type, sampling

methodology, sample size, and generalizability previously discussed. Comparisons drawn within the set of CLD evaluations for the study are particularly prone to issues of sample size due to the relatively small size of the subsample (n = 50). Generalizability is further limited by the ethnic and linguistic composition of the CLD group, with the majority of reports written on CLD students from Hispanic backgrounds whose linguistic history was characterized by Spanish. Again, future research with a larger sample of evaluations completed in multiple districts across the nation would promote greater generalizability. Moreover, given that laws on bilingual education and educational practices pertaining to CLD students can vary significantly across states, replicating this research in other areas would allow for useful comparisons to be drawn in assessment trends.

Finally, further research is needed to better understand the role of RTI in psychoeducational evaluation practices with CLD students. The present study did not find evidence of RTI having an increased role in the evaluation process for CLD students when compared with non-CLD students. It is possible that the role of RTI in helping psychologists rule out linguistic proficiency as the chief cause of learning difficulties may simply not be discernible in psychoeducational evaluation reports. Alternatively, RTI's greatest potential may lie in remediating learning difficulties in CLD students, thereby preventing them from ever needing to be evaluated; research has provided initial support for this (Kamps et al., 2007; Vaughn, Linan-Thompson et al., 2006; Vaughn, Mathes et al., 2006). As time continues to pass since IDEIA's (2004) endorsement of RTI, large-scale studies on eligibility decision-making will help determine whether many

practitioners are formally using failure to respond to intervention to qualify students for special education, and if so, whether the practice is more common with CLD students. At the same time, more qualitative studies with practitioners on how they use information on interventions attempted and data on progress to determine that linguistic proficiency is not the cause of academic issues will be useful.

Study Summary

The present study examined the integration of RTI data in the psychoeducational evaluation process for students found eligible for special education services under the SLD category. Psychoeducational evaluation reports were sampled from a school district that has been incorporating an RTI framework in its schools since 2005. A total of 212 psychoeducational reports authored by 25 certified school psychologists were reviewed. A subsample of 50 reports written on students from culturally and linguistically diverse backgrounds was also analyzed.

One major research area addressed the integration of RTI data and traditional assessment practices with all students. Although recently completed reports were more likely to reference the prereferral intervention process than older reports, they were no more likely to contain RTI data or cite empirically-based interventions than older reports. RTI data was present in a minority of reports in the sample, although it was more common at the elementary level than at the secondary level, and more likely to focus on reading than math or writing. There was no relationship found between the amount of traditional testing performed and the tendency to reference prereferral strategies and interventions. However, reports that included RTI data tended to include more cognitive

testing than reports without this data. Overall, though, it was exceedingly rare for practitioners to use failure respond to intervention to formally determine SLD eligibility, despite the district being permitted to do so by the state. The vast majority of practitioners still employed a discrepancy model for eligibility determination purposes, even if the RTI process was discussed and data was included. These results indicate that there is great variability with regard to the way RTI processes and procedures are documented in psychoeducational evaluation reports. Future research should expand the study of this topic to additional districts and regions, as well as employ qualitative methods with practitioners to better understand how they use RTI documentation to make decisions.

A second area of focus was assessment practices with CLD students, including, but not limited to, the role of RTI data in making special education eligibility recommendations. Evaluations of CLD students did not include a greater variety of traditional assessments than evaluations of non-CLD students, but they were more likely to include and rely solely upon nonverbal cognitive measures; the latter was true regardless of whether the practitioner was bilingual. Even so, full cognitive assessments and academic achievement measures administered in English were still the most commonly used traditional measures. Results did not support the idea that practitioners documented prereferral efforts, cited data on response to targeted interventions, or included data on progress within the general curriculum to a greater degree for CLD students than non-CLD students. However, most practitioners did incorporate data from individually-administered language proficiency tests. These findings indicate that current

practices with CLD students may not always match best practice recommendations. Further research is needed to confirm these trends beyond the district of focus.

In sum, the present study provided a look at current practices in SLD identification through the lens of psychoeducational evaluation reports. The results illustrated the gap between research and practice that exists with regard to RTI, as well as highlighted similarities and differences in evaluation practices for CLD and non-CLD students suspected of having learning disabilities.

REFERENCES

Acevedo-Polakovich, I., Reynaga-Abiko, G., Garriott, P., Derefinko, K., Wimsatt, M., Gudonis, L.C., & Brown, T. L. (2007). Beyond instrument selection: Cultural considerations in the psychological assessment U.S. Latinas/os. *Professional Psychology: Research and Practice, 38,* 375-384. doi:10.1037/0735-7028.38.4.375

Al Otaiba, S., Petscher, Y., Pappamihiel, N. E., Williams, R. S., Dyrlund, A. K., & Connor, C. (2009). Modeling oral reading fluency development in Latino students: A longitudinal study across second and third grade. *Journal of Educational Psychology, 10,* 315-329. doi:10.1037/a0014698

Amato, J., & Watkins, M. W. (2011). The predictive validity of CBM writing indices for eighth-grade students. *Journal of Special Education, 44*, 195-204. doi:10.1177/0022466909333516

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57,* 1060-1073. doi:10.1037//0003-066X.57.12.1060

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

Ardoin, S. P., & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Estimates of standard error when monitoring progress using alternate passage sets. *School Psychology Review, 38,* 266-283.

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

Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning, decision making, and the identification of children in need of services. *Journal of Psychoeducational Assessment, 23,* 362-380. doi:10.1177/073428290502300405

Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L., . . . Williams, K. L. (2004). Examining the incremental benefits of administering a

maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review, 33,* 218-233.

Arizona Revised Statutes (ARS), Title 15: Education, Chapter 7: Instruction, (2012).

Argulewicz, E. N. (1983). Effects of ethnic membership, socioeconomic status, and home language on LD, EMR, and EH placements. *Learning Disabilities Quarterly, 6,* 195-200. doi:10.2307/1510797

Arizona Department of Education. (2009a). *Arizona monitoring system*. Retrieved from https://www.azed.gov/ess/programsupport/Monitoring/2009-2010MonitoringManual.pdf

Arizona Department of Education. (2009b). *Primary home language other than English (PHLOTE) home language survey*. Phoenix, AZ: Author.

Arizona Department of Education. (2011). *Primary home language other than English (PHLOTE) home language survey*. Phoenix, AZ: Author.

Arizona Department of Education. (2012a). *Percentage of free and reduced reports*. Retrieved from http://www.azed.gov/health-nutrition/frpercentages/

Arizona Department of Education. (2012b). RTI glossary. Retrieved from http://www.azed.gov/wp-content/uploads/PDF/RTIGLOSSARY.pdf

Arizona Department of Education, & Harcourt Assessment. (2007). *AZELLA technical manual*. Phoenix, AZ: Arizona Department of Education.

Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children, 71*, 283-300. doi:10.1177/001440290507100305

August, D., & Shanahan, T. (2006). *Developing literacy in second language learners: Report of the national literacy panel on language-minority children and youth.* Mahwah, NJ: Erlbaum.

Bailey, A. L., & Huang, B. H. (2011). Do current English language development/proficiency standards reflect the English needed for success in school? *Language Testing, 28,* 343-365. doi:10.1177/0265532211404187

Bailey, A. L., & Kelly, K. R. (2010). *The use and validity of home language surveys in state English language proficiency assessment systems: A review and issues perspective*. Los Angeles, CA: University of California Los Angeles.

Bailey, A. L., & Kelly, K. R. (2012). Home language survey practices in the initial identification of English learners in the United States. *Educational Policy, 20,* 1-35. doi:10.1177/0895904811432137

Bainter, T.R., & Tollefson, N. (2003). Intellectual assessment of language minority students: What do school psychologists believe are acceptable practices? *Psychology in the Schools, 40,* 899-903. doi:10.1002/pits.10131

Bartholomew, J. (2004). *Measuring intelligence: Facts and fallacies.* New York, NY: Cambridge University Press.

Bayat, M., Mindes, G., & Covitt, S. (2010). What does RTI (response to intervention) look like in preschool? *Early Childhood Education Journal, 37,* 493-500. doi:10.1007/s10643-010-0372-6

Berkeley, S., Bender, W. N., Peaster, L. G., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities, 19,* 579-586. doi:10.1177/0022219408326214

Berninger, V., Vaughan, K., Abbott, R., Brooks, A., Begay, K., Curtin, G. . . . Graham, S. (2000). Language-based spelling instruction: Teaching children to make multiple connections between spoken and written words. *Learning Disability Quarterly, 23,* 117-135. doi:10.2307/1511141

Berninger, V. W., Winn, W. D., Stock, P., Abbott, R. D. Eschen, K., Lin, S., … Nagy, W. (2008). Tier 3 specialized writing instruction for students with dyslexia. *Reading and Writing, 21,* 95-129. doi:10.1007/s11145-007-9066-x

Blatchley, L A., & Lau, M. Y. (2010). Culturally competent assessment of English language learners for special education services. *Communiqué, 38,* 1-8.

Bouman, S. H. (2010). *Response-to-intervention in California public schools: Has it helped address disproportional placement rates for students with learning disabilities?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses Database. (UMI No. 3414045)

Brown-Chidsey, R., Bronaugh, L., & McGraw, K. (2009). *RTI in the classroom: Guidelines and recipes for success*. New York, NY: The Guilford Press.

Burns, M. K., & Coolong-Chaffin, M. (2006). Response to intervention: The role of and effect on school psychology. *School Psychology Forum, 1,* 3-15.

Burns, M. K., Vanderwood, M., & Ruby, S. (2005). Evaluating the readiness of prereferral intervention teams for use in a problem-solving model: Review of three levels of research. *School Psychology Quarterly, 20,* 89-105.

Campbell, H., Espin, C. A,. & McMaster, K. (2013). The technical adequacy of CBM writing measures with English language learners. *Reading and Writing, 26,* 431-452.

Carrasquillo, A. L., & Rodriguez, J. (1997). Hispanic limited English-proficient students with disabilities: A case study example. *Learning Disabilities: A Multidisciplinary Journal, 8,* 167-174.

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

Christ, T. J., Scullin, S., Tolbize, A., & Jiban, C. L. (2008). Implications of recent research: Curriculum based measurement of math computation. *Assessment for Effective Intervention, 33,* 198-205. doi:10.1177/1534508407313480

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

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

Cirino, P. T., Vaughn, S., Linan-Thompson, S., Cardenas-Hagan, E., Fletcher, J. M., & Francis, D. J. (2009). One-year follow-up outcomes of Spanish and English interventions for English language learners at risk for reading problems. *American Educational Research Journal, 46,* 744-781. doi:10.3102/0002831208330214

Civil Rights Act of 1964, Pub. L. 88-352 (1964).

Clarke, B., Smolkowski, K., Baker, S. K., Fien, H., Doabler, C., & Chard, D. (2011). The impact of a comprehensive tier 1 kindergarten curriculum on the achievement of students at-risk in mathematics. *The Elementary School Journal, 111,* 561-584.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46. doi:10.1177/001316446002000104

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159. doi:10.1037//0033-2909.112.1.155

Cohen, S., Sullivan, A. L., & Caterino, L. C. (2009, February). *Using RTI data to diagnose SLD in culturally and linguistically diverse student*. Paper presentation at the annual conference of the National Association of School Psychologists, Boston, MA.

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

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

Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., & Hamlett, C. L. (2012). The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities, 45,* 79-95. doi:10.1177/0022219410393012

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

Cornwall, A. (1990). Social validation of psychoeducational assessment reports. *Journal of Learning Disabilities*, *23*, 413-416. doi:10.1177/002221949002300704

Coutinho, M. J., & Oswald, D. P. (2000). Disproportionate representation in special education: A synthesis and recommendations. *Journal of Child and Family Studies, 9*, 135-156.

Curtis, M. J., Lopez, A. D., Castillo, J. M., Batsche, G. M., Minch, D., & Smith, J. C. (2008). The status of school psychology: Demographic characteristics, employment conditions, professional practices, and continuing professional development. *Communiqué, 36,* 27-29.

Daly, E. J., III, Wright, J. A., Kelly, S. Q., & Martens, B. K. (1997). Measures of early academic skills: Reliability and validity with a first grade sample. *School Psychology Quarterly, 12,* 268-280. doi:10.1037/h0088962

Dean, V. J., Burns, M. K., Grialou, T., & Varro, P. J. (2006). Comparison of ecological validity of learning disabilities diagnostic model. *Psychology in the Schools, 43*, 157-168. doi:10.1002/pits.20143

DeAvila, A., & Duncan, S. E. (1990). *Language Assessment Scales oral technical report*. Monterey, CA: Macmillan-McGraw-Hill.

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

Deno, S. L. (2002). Problem solving as "best practice." In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., Vol. 1, pp. 37–55). Bethesda, MD: National Association of School Psychologists.

Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*, 184-192. doi:10.1177/00224669030370030801

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*, 1-16.

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*, 507-526.

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

Diercks-Gransee, B., Weissenburger, J. W., Johnson, C. L., & Christensen, P. (2009). Curriculum-based measures of writing for high school students. *Remedial and Special Education, 30*, 360-371. doi:10.1177/0741932508324398

Dombrowski, S. C., Kamphaus, R. W., Barry, M., Brueggeman, A., Cavanagh, S., Devine, K., . . . Vess, S. (2006). The Solomon Effect in learning disabilities diagnosis: Have we not yet learned from history? *School Psychology Quarterly, 21*, 359-373.

Donovan, S., & Cross, C. (2002). *Minority students in special and gifted education.* Washington, DC: National Academy Press.

Duhon, G. J., House, S. H., & Stinnett, T. A. (2012). Evaluating the generalization of math fact fluency gains across paper and computer performance modalities. *Journal of School Psychology, 50*, 335-345. doi:10.1016/j.jsp.2012.01.003

Duhon, G. J., Noell, G. H., Witt, J. C., Freeland, J. T., Dufrene, B. A., & Gilbertson, D. N. (2004). Identifying academic skill and performance deficits: The experimental analysis of brief assessments of academic skills. *School Psychology Review, 33*, 429-443.

Dulaney, S. K. (2013). A middle school's response-to-intervention journey: Building systematic processes of facilitation, collaboration, and implementation. *NASSP Bulletin, 97*, 53-77. doi:10.1177/0192636512469289

Education of All Handicapped Children Act of 1975, Pub. L. No. 94-142 (1975).

Espin, C., & Tindal, G. (1998). Curriculum-based measurement for secondary students. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 214-253). New York, NY: Guilford.

Espin, C. A., Wallace, T., Campbell, H., Lembke, E., Long, J., & Tichá, R. (2008). Curriculum-based measurement in writing: Predicting the success of high-school students on state standards tests. *Exceptional Children, 74,* 174-193. doi:10.1177/001440290807400203

Evans, J.J., Floyd, R.G., McGrew, K.S., & Leforgee, M.H. (2002). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review, 31,* 364-372.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149-1160. doi:10.3758/BRM.41.4.1149

Feuerstein, R. (1979). *The dynamic assessment of retarded individuals: The Learning Potential Assessment Device, theory, instruments, and techniques.* Baltimore, MD: University Park.

Fewster, S., & MacMillan, P. D. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial Special Education, 23,* 149-156. doi:10.1177/07419325020230030301

Figueroa, R. A., & Sassenrath, J. M. (1989). A longitudinal study of the predictive validity of the System of Multicultural Pluralistic Assessment (SOMPA). *Psychology in the Schools, 26,* 5-19. doi:10.1002/1520-6807(198901)26:1%3C5::AID-PITS2310260102%3E3.0.CO;2-D

Fisher, D., & Frey, N. (2011). Implementing RtI in a high school: A case study. *Journal of Learning Disabilities, 46,* 99-114. doi:10.1177/0022219411407923

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (Eds.). (2013). *Essentials of cross-battery assessment* (3rd ed.). New York, NY: John Wiley.

Florida Center for Reading Research. (2014). *Florida Center for Reading Research.* Retrieved from http://www.fcrr.org/

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across school-age years. *Psychology in the Schools, 40,* 155-171. doi:10.1002/pits.10083

Foegen, A., Jiban, C., & Deno, S. L. (2007). Progress monitoring in mathematics: A review of the literature. *Journal of Special Education, 41*, 121-139. doi:10.1177/00224669070410020101

Fry, R. (2007). *How far behind in math and reading are English language learners?* Washington, DC: Pew Hispanic Center.

Fry, R. (2008). *The role of schools in the English language learner achievement gap.* Washington, DC: Pew Hispanic Center.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, N. C. (2008). Making "secondary intervention" work in a three tier responsiveness-to-intervention model: Findings from the first grade longitudinal reading study of the National Research Center on Learning Disabilities. *Reading and Writing Quarterly: An Interdisciplinary Journal, 21,* 413-436. doi:10.1007/s11145-007-9083-9

Fuchs, D., Compton, D.L., Fuchs, L.S., Bryant, J., Hamlett, C., & Lambert, W. (2012). First-grade cognitive abilities as long-term predictors of reading comprehension and disability status. *Journal of Learning Disabilities, 45*, 217-231. doi:10.1177/0022219412442154

Fuchs, L. S., Compton, D. C., Fuchs, D., Hollenbeck, K., Hamlett, C. L., & Seethaler, P. M. (2011). Two-stage screening for math word-problem difficulty using dynamic assessment of algebraic learning. *Journal of Learning Disabilities, 44*, 372-380.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 58*, 45-58.

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 responsiveness-to-intervention: What, why, and how valid is it? *Reading Research Quarterly, 41,* 92-99. doi:10.1598/RRQ.41.1.4

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*, 7-21. doi:10.1177/001440290407100101

Fuchs, D., Fuchs, L. S., & Compton, D. (2012). SMART RTI: Next generation approach to multilevel prevention. *Exceptional Children, 78,* 263-279. doi:10.1177/001440291207800301

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal measures of reading comprehension. *Remedial and Special Education, 9,* 20-28. doi:10.1177/074193258800900206

Fuchs, L.S., Fuchs, D., & Stecker, P. (2010). The "blurring" of special education in a new continuum of general education placements and services. *Exceptional Children, 76,* 301-323. doi:10.1177/001440291007600304

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Waltz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22,* 27-48.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Response-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice, 18,* 157-171. doi:10.1111/1540-5826.00072

Garcia, E., Lawton, K., & Diniz de Figueiredo, E. H. (2010). A*ssessment of young English language learners in Arizona: Questioning the validity of the state measure of English proficiency*. Los Angeles, CA: The Civil Rights Project.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114,* 345-362. doi:10.1037//0033-2909.114.2.345

George, M. C. (1999). A new IDEA: The individuals with disabilities act after 1997 amendments. *Law and Psychology Review, 23,* 91-130.

Gersten, R., Compton, D., Connor, C. M., Dimíno. J., Santoro, L, Linan-Thompson. S.. & Tilly, D. W. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention in the primary grades.* Washington, DC: Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance.

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *Journal of Special Education, 40,* 103-114. doi:10.1177/00224669060400020101

Goldenberg, C. (2007). Improving achievement for English learners: Conclusions from recent reviews and emerging research. In G. Li & P.A. Edwards (Eds.), *Best practices in ELL instruction* (pp. 15-43). New York, NY: Guilford.

Goldenberg, C. (2008). Teaching English language learners: What the research does and does not say. *American Educator, 33,* 8-23.

Goldenberg, C. & Rutherford-Quach, S. (2012). The Arizona home language survey: The under-identification of students for English language services. *Language Policy, 11,* 21-30. doi:10.1007/s10993-011-9224-5

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Good, R. H., & Shinn, M. R. (1990). Forecasting accuracy of slope estimates for reading curriculum-based measurement: Empirical evidence. *Behavioral Assessment, 12,* 179–193.

Graney, S. B., Martinez, R. A., Missall, K. N., & Aricak, 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

Graney, S. B., Missall, K., Martinez, M., & Bergstrom, M. (2009). A preliminary investigation of within-year growth patterns in reading and mathematics curriculum-based measures. *Journal of School Psychology, 47,* 121-142. doi:10.1016/j.jsp.2008.12.001

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124,* 75-111. doi:10.1037//0033-2909.124.1.75

Hale, J. B., 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 and intervention: An expert white paper consensus. *Learning Disability Quarterly, 33,* 223-236.

Hale, J. B., Fiorello, C. A., Miller, J. A., Wenrich, K., Teodori, A., & Henzel, J. N. (2008). WISC-IV interpretation for specific learning disabilities and intervention: A cognitive hypothesis testing approach. In A. Prifitera, D. H. Saklofske, & E. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (2nd ed., pp. 109-171). New York, NY: Elsevier.

Hale, J. B., Kaufman, A., Naglieri, J. A., & Kavale, K. A. (2006). Implementation of IDEA: Integrating response to intervention and cognitive assessment methods. *Psychology in the Schools, 43,* 753-770. doi:10.1002/pits.20186

Hale, J. B., Naglieri, J. A., Kaufman, A. S., & Kavale, K. A. (2004). Specific learning disability classification in the new Individuals with Disabilities in Education Act: The danger of good ideas. *The School Psychologist, 1,* 6-29.

Hallahan, D. P., & Mercer, C. D. (2002). Learning disabilities: Historical perspectives. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification in learning disabilities: Research to practice* (pp. 1-67). Mahwah, NJ: Erlbaum.

Harcourt Assessment. (2004). *Stanford English language proficiency test*. San Antonio, TX: Author.

Hart, J. (2009). Strategies for culturally and linguistically diverse students with special needs. *Preventing School Failure, 53,* 197-208. doi:10.3200/PSFL.53.3.197-208

Harvey, V. S. (1997). Improving readability of psychological reports. *Professional Psychology: Research and Practice, 28,* 271-274. doi:10.1037//0735-7028.28.3.271

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

Hernández Finch, M. E. (2012). Special considerations with response to intervention and instruction for students with diverse backgrounds. *Psychology in the Schools*, *49,* 285-296. doi:10.1002/pits.21597

Hintze, J., Shapiro, E., Conte, K., & Basile, I. (1997). Oral reading fluency and authentic reading material: Criterion validity of the technical features of CBM survey-level assessment. *School Psychology Review, 26,* 535-553.

Honigsfeld, A. (2009). ESL programs: Not "One size fits all." *KDP Record, 45*(4), 166-171.

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review, 34,* 9-26.

Hosp, M. K., & Hosp, J. (2003). Curriculum-based measurement for reading, math, and spelling: How to do it and why. *Preventing School Failure, 48,* 10-17. doi:10.1080/1045988X.2003.10871074

Hosp, J. L., & Reschly, D. J. (2004). Disproportionate representation of minority students in special education: Academic, economic, and demographic predictors. *Exceptional Children, 70,* 185-200. doi:10.1177/001440290407000204

Ikeda, M. J., Neessen, E., & Witt, J. C. (2008). Best practices in universal screening. In A. Thomas & J. Grimes (Eds.) *Best practices in school psychology* (5th ed., Vol. 2., pp. 103-114). Bethesda, MD: National Association of School Psychologists.

Individuals with Disabilities Education Act of 1990, 20 U.S.C. § 1400 *et seq.* (1990).

Individuals with Disabilities Education Act of 1997, 20 U.S.C. § 1400 *et seq.* (1997).

Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. § 1400 *et seq.* (2004).

Jacob, S., & Hartshorne, T. S. (2007). *Ethics and law for school psychologists* (5th ed.). Hoboken, NJ: John Wiley & Sons.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for service delivery in an RTI framework: Candidate measures. *School Psychology Review, 36,* 582-599.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children, 59,* 421-432.

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,* 36-46. doi:10.1177/0731948712464963

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review, 30,* 420-437.

Johnson, E. S., Humphrey, M., Mellard, D. F., Woods, K., & Swanson, H. L. (2010). Cognitive processing deficits and students with specific learning disabilities: A selective meta-analysis of the literature. *Learning Disability Quarterly, 33,* 3-18.

Joseph, L. M. (2008). Best practices on interventions for students with reading problems. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., Vol. 2, pp. 1163-1180). Bethesda, MD: National Association of School Psychologists.

Kamps, D., Abbott, M., Greenwood, C., Arreaga-Mayer, C., Wills, H., Longstaff, J., et al. (2007). Use of evidence-based, small group reading instruction for English language learners in elementary grades: Secondary-tier intervention. *Learning Disability Quarterly, 30,* 153-168. doi:10.2307/30035561

Kanaya, T., & Ceci, S. J. (2012). The impact of the Flynn effect on LD diagnoses in special education. *Journal of Learning Disabilities, 45,* 319-326. doi:10.1177/0022219410392044

Kavale, K. A. (2002). Discrepancy models in the identification of learning disabilities. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 370-371). Mahwah, NJ: Erlbaum.

Kavale, K. A., Holdnack, J. A., & Mostert, M. P. (2005). Responsiveness to intervention and the identification of specific learning disability: A critique and alternative proposal. *Learning Disability Quarterly, 28,* 2-16. doi:10.2307/4126970

Keller-Margulis, M. A., Payan, A., & Booth, C. (2012). Reading curriculum-based measures in Spanish: An examination of validity and diagnostic accuracy. *Assessment for Effective Intervention, 37,* 212-223. doi:10.1177/1534508411435721

Kirk, S. A. (1962). *Educating exceptional children.* Boston, MA: Houghton Mifflin.

Kirk, S. A. (1963). Behavioral diagnosis and remediation of learning disabilities. In Anonymous, *Proceedings of the conference on exploration into problems of the perceptually handicapped child.* Chicago, IL: Perceptually Handicapped Children.

Kirk, S. A. (1977). Specific learning disabilities. *Journal of Clinical Child Psychology, 6,* 23-26. doi:10.1080/15374417709532776

Klingner, J. K., Artiles, A. J., & Mendez Barletta, L. (2006). English Language Learners who struggle with reading: Language acquisition or LD? *Journal of Leaming Disabilities, 39,* 108-128. doi:10.1177/00222194060390020101

Klingner, J. K., & Harry, B. (2006). The special education referral and decision-making process for English Language Learners: Child study team meetings and staffings. *Teachers College Record, 108,* 2247-2281.

Kozleski, E. B. & Huber, J. J., (2010). Systemic change for RTI: Embedding change within a critical framework. *Theory into Practice, 49,* 258-264.

Lakin, J. M. (2012). Assessing the cognitive abilities of culturally and linguistically diverse students: Predictive validity of verbal, quantitative, and nonverbal tests. *Psychology in the Schools, 49,* 756-768. doi:10.1002/pits.21630

Lawton, S. B. (2012). State education policy formation: The case of Arizona's English language learner legislation. *American Journal of Education, 118,* 455-487. doi:10.1086/666361

Lembke, E., Hampton, D., & Beyers, S. (2012). Response to intervention in mathematics: Critical elements. *Psychology in the Schools, 49,* 257-272. doi:10.1002/pits.21596

Linan-Thompson, S., Vaughn, S., Prater, K., & Cirino, P. (2006). The response to intervention of English language learners at risk for reading problems. *Journal of Learning Disabilities, 39,* 390-398. doi:10.1177/00222194060390050201

Logan, J. A. R., & Petscher, Y. (2010). School profiles of at-risk student concentration: Differential growth in oral reading fluency. *Journal of School Psychology, 48*(2), 163-186. doi:10.1016/j.jsp.2009.12.002

Lopez, E. C., & Bursztyn, A. M. (2013). Future challenges and opportunities: Toward culturally responsive training in school psychology. *Psychology in the Schools, 50,* 212-228. doi:10.1002/pits.21674

Lopez, F. A., & Thompson, S. S. (2011). The relationship among measures of written expression using curriculum-based measurement and the Arizona instrument to

measure skills (AIMS) at the middle school level. *Reading and Writing Quarterly, 27*, 129-152. doi:10.1080/10573561003769640

MacSwan, J., & Rolstad, K. (2006). How language proficiency tests mislead us about ability: Implications for English language learner placement in special education. *Teachers College Record*. *108*, 2304-2328. doi:10.1111/j.1467-9620.2006.00783.x

Mahoney, K., Haladyna, T., & MacSwan, J. (2009). The need for multiple measures in reclassification decisions: A validity study of the Stanford English Language Proficiency Test (SELP). In J. S. Lee, T. G. Wiley, & R. Rumberger (Eds.), *The Education of Language Minority Immigrants in the USA* (pp. 263-294). Bristol, UK: Multilingual Matters.

Mallin, B., Schellenberg, M. E., & Smith, C. (2012). A modified content analysis of existing school psychology reports. *Canadian Journal of School Psychology, 27,* 243-257. doi:10.1177/0829573512449997

Markell, M. A., & Deno, S.L. (1997). Effects of increasing oral reading: Generalization across reading tasks. *Journal of Special Education, 31*, 233-250. doi:10.1177/002246699703100205

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based* measurement: Assessing special children (pp. 18–78). New York, NY: Guilford.

Marston, D. B. (2005). Tiers of intervention in responsiveness to intervention: Prevention outcomes and learning disabilities identification patterns. *Journal of Learning Disabilities, 38*, 539-544. doi:10.1177/00222194050380061001

McCook, J. E. (2006). *The RTI guide: Developing and implementing a model in your schools.* Horsham, PA: LRP Publications.

McDonald, J. H. (2009). *Handbook of biological statistics* (2nd ed.). Baltimore, MD: Sparky House.

McMaster, K., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in writing. *Journal of Special Education, 41*, 68-84.

McMaster, K. L., Wayman, M., & Cao, M. (2006). Monitoring the reading progress of secondary-level English learners: Technical features of oral reading and maze tasks. *Assessment for Effective Intervention, 31*, 17-32. doi:10.1177/073724770603100402

Mellard, D. F., McKnight, M., & Jordan, J. (2010). RTI tier structures and instructional intensity. *Learning Disabilities Research and Practice, 25*, 217-225. doi:10.1111/j.1540-5826.2010.00319.x

Miller, J. A., & Watkins, M. W. (2010). The use of graphs to communicate psychoeducational test results to parents. *Journal of Applied School Psychology, 26,* 1-16. doi:10.1080/15377900903175911

Morris, J. (1977). What tests do schools use with Spanish-speaking students? *Integrated Education, 15*, 21-37. doi:10.1080/0020486770150207

Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P. . . . Shaywitz, B. A. (1998). Subtypes of reading disability: Coherent variability around a phonological core. *Journal of Educational Psychology, 90*, 347-373.

Muñoz-Sandoval, A. F., Cummins, J., Alvarado, C. G., & Ruef, M. L. (2005). *Bilingual Verbal Ability Tests – Normative Update*. Itasca, IL: Riverside Publishing.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System. Administration and scoring manual.* Rolling Meadows, IL: Riverside Publishing.

National Center on Response to Intervention. (2010). *Essential components of RTI: A closer look at response to intervention.* Washington, DC: Author.

National Institute of Child Health and Human Development. (2000). Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769). Washington, DC: U.S. Department of Education.

National Research Council. (2011). *Allocating federal funds for state programs for English language learners*. Washington, DC: National Academies Press.

Nellis, L. M. (2012). Maximizing the effectiveness of building teams in response to intervention implementation. *Psychology in the Schools, 49*, 245-256. doi:10.1002/pits.21594

New America Foundation. (2013). *Federal education budget project: Comparative analysis of funding, student demographics and achievement data.* Retrieved from http://febp.newamerica.net/k12

No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2001).

Noell, G. H., Freeland, J. T., & Witt, J. C. (2001). Using brief assessments to identify effective interventions for individual students. *Journal of School Psychology, 39*, 335-355. doi:10.1016/S0022-4405(01)00072-3

Noell, G. H., Gansle, K. A., Witt, J. C., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. N., . . . Northup, J. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. *Journal of Applied Behavior Analysis, 31,* 659-663. doi:10.1901/jaba.1998.31-659

Nuttall, E. V. (1987). Survey of current practices in the psychological assessment of limited-English-proficiency handicapped children. *Journal of School Psychology, 25,* 53-61. doi:10.1016/0022-4405(87)90060-4

O'Bryon, E. C., & Rogers, M. R. (2010). Bilingual school psychologists' assessment practices with English language learners. *Psychology in the Schools, 47,* 1018-1034. doi:10.1002/pits.20521

O'Connor, R. E. (2000). Increasing the intensity of intervention in kindergarten and first grade. *Learning Disabilities Research and Practice, 15,* 43-54. doi:10.1207/SLDRP1501_5

Ochoa, S. H., González, D., Galarza, A., & Guillemard, L. (1996). The training and use of interpreters in bilingual psychoeducational assessment: An alternative in need of study. *Diagnostique, 21,* 19-22.

Ochoa, S. H., Powell, M. P., & Robles-Piña, R. (1996). School psychologists' assessment practices with bilingual and limited- English-proficient students. *Journal of Psychoeducational Assessment, 14,* 250-275. doi:10.1177/073428299601400306

Ochoa, S. H., Rivera, B. D., & Powell, M. P. (1997). Factors used to comply with the exclusionary clause with bilingual and limited-English-proficient pupils: Initial guidelines. *Learning Disabilities Research & Practice, 12,* 161-167.

Olvera, P., & Gomez-Cerrillo, L. (2011). A bilingual approach (English & Spanish) psychoeducational assessment model grounded in Cattell-Horn-Carroll (CHC) theory: A cross battery approach. *Contemporary School Psychology, 15,* 113-123.

Orosco, M. J. & Klingner, J. (2010). One school's implementation of RTI with English Language Learners: "Referring into RTI". *Journal of Learning Disabilities, 43,* 269-288.

Ortiz, S. (2004). *Comprehensive assessment of culturally and linguistically diverse students: a systematic, practical approach for nondiscriminatory assessment.* Retrieved from http://www.nasponline.org/resources/culturalcompetence/ortiz.pdf

Orton, S. T. (1925). "Word-blindness" in school children. *Archives of Neurology and Psychiatry, 14,* 581-615. doi:10.1001/archneurpsyc.1925.02200170002001

Parker, R. I., Tindal, G., & Hasbrouck, J. (1991). Progress monitoring with objective measures of writing performance for students with mild disabilities. *Exceptional Children, 58,* 61-73.

Parker, R., & Tindal, G. (1992). Estimating trend in progress monitoring data: A comparison of simple line-fitting methods. *School Psychology Review, 21,* 300-312.

Parkin, J., & Beaujean, A. A. (2012). The effects of Wechsler Intelligence Scale for Children-Fourth Edition cognitive abilities on math achievement. *Journal of School Psychology, 50,* 113-128. doi:10.1016/j.jsp.2011.08.003

Pray, L. (2005). How well do commonly used language instruments measure English oral language proficiency? *Bilingual Research Journal. 29,* 387-409. doi:10.1080/15235882.2005.10162841

Prewett, S., Mellard, D.F., Deshler, D.D., Allen, J., Alexander, R. & Stern, A. (2012). Response to intervention in middle schools: Practices and outcomes. *Learning Disabilities Research and Practice, 27,* 136-147. doi:10.1111/j.1540-5826.2012.00359.x

Pryzwansky, W. B., & Hanania, J. S. (1986). Applying problem-solving approaches to school psychological reports. *Journal of School Psychology, 24,* 133-141. doi:10.1016/0022-4405(86)90005-1

Ragan, A., & Lesaux, N. (2006). Federal, state, and district level English language learner program entryand exit requirements: Effects on the education of language minority learners. *Education Policy Analysis Archives, 14*(20). Retrieved from http://epaa.asu.edu/epaa/v14n20

Ramírez, R. D., & Shapiro, E. S. (2007). Cross-language relationship between Spanish and English oral reading fluency among Spanish-speaking English language learners in bilingual education classrooms. *Psychology in the Schools, 44,* 795-806.

Reschly, A., Busch, T., Betts, J., Deno, S., & Long, J. (2009). Curriculum-based measurement of oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47,* 427-469. doi:10.1016/j.jsp.2009.07.001

Reynolds, C. R., & Shaywitz, S. E. (2009). Response to intervention: Prevention and remediation, perhaps. Diagnosis, no. *Child Development Perspectives, 3,* 44-47. doi:10.1111/j.1750-8606.2008.00075.x

Rolstad, K., Mahoney, K. S., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy, 19*, 572-594. doi:10.1177/0895904805278067

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16*, 181-197. doi:10.1111/j.1469-7610.1975.tb01269.x

Samson, J.& Lesaux, N. K. (2009). Language minority learners in special education: Rates and predictors of identification for services. *Journal of Learning Disabilities, 42*, 148-162. doi:10.1177/0022219408326221

Schwartz, N. H., & Wilkinson, W. K. (1987). Perceptual influence of psychological reports. *Psychology in the Schools, 24*, 127-135.

Scott, A. N., Boynton Hauerwas, L., & Brown, R. D. (2014). State policy and guidance for identifying learning disabilities in culturally and linguistically diverse students. *Learning Disability Quarterly, 37*, 172-185. doi:10.1177/0731948713507261

Seethaler, P. M., & Fuchs, L. S. (2010). The predictive utility of kindergarten screening for math difficulty. *Exceptional Children, 77*, 37-59.

Seethaler, P. M., & Fuchs, L. S. (2011). Using curriculum-based measurement to monitor kindergarteners' mathematics development. *Assessment for Effective Intervention, 36*, 219-229. doi:10.1177/1534508411413566

Shifrer, D., Muller, C., & Callahan, R. M. (2011). Disproportionality and learning disabilities: Parsing apart race, socioeconomic status, and language. *Journal of Learning Disabilities, 44*, 246-257. doi:10.1177/0022219410374236

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

Shinn, M. R. (2002). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., Vol. 1, pp. 671-697). Bethesda, MD: NASP Publications.

Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within RTI: Research on educational need and benefit from academic intervention. *School Psychology Review, 36*, 601-617.

Shinn, M. R., Good, R. H., III, & Stein, S. (1989). Summarizing trend in student achievement: A comparison of methods. *School Psychology Review, 18*, 356-370.

Shores, C., & Chester, K. (2009). *Using RTI for school improvement: Raising every student's achievement scores*. Thousand Oaks, CA: Corwin Press.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York, NY: McGraw-Hill.

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research, 75*, 247-284. doi:10.3102/00346543075002247

Skirbekk, V.,Stonawski, M.,Bonsang, E., &Staudinger, U. M. (2013). The Flynn effect and population aging. *Intelligence, 41*, 169-177. doi:10.1016/j.intell.2013.02.001

Speece, D. L., Case, L. P., & Molloy, D. E. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research & Practice, 18*, 147-156. doi:10.1111/1540-5826.00071

Sternberg, R. J. & Grigorenko, E. L. (1999). *Our labeled children: What every parent and teachersneeds to know about learning disabilities*. Reading, MA: Perseus Books.

Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage bypracticing school psychologists: A national survey. *Journal of Psychoeducational Assessment, 12*, 331-350.

Stuebing, K. K., Barth, A., Molfese, P., Weiss, B., & Fletcher, J.(2009). IQ is not strongly related to response to reading instruction; A meta-analytic interpretation. *Exceptional Children, 76*, 31-51.

Stuebing, K. K., Fletcher, J. M., Francis, D. J., & Branum-Martin, L. (2012). Evaluation of the technical adequacy of three methods for identifying specific learning disabilities based on cognitive discrepancies. *School Psychology Review, 41*, 3-22.

Stuebing, K. K., Fletcher, J. M., LeDoux, J., Lyon, G. R., Shaywitz, S., & Shaywitz,B. (2002). Validity of IQ discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal, 39*, 469-518. doi:10.3102/00028312039002469

Styck, K. M., Watkins, M. W., & Vanderwood, M. (2013). Diagnostic utility of the Culture-Language Interpretive Matrix for the Wechsler Intelligence Scales for Children - Fourth Edition among referred students. *School Psychology Review, 42*, 367-382.

Sullivan, A. L. (2011). Disproportionality in special education identification and placement of English language learners. *Exceptional Children, 77*, 317-334.

doi:10.1177/001440291107700304

Sullivan, A. L., & Cohen, S. (2012, August). *Integration of traditional assessment and RTI in evaluations of learning disabilities.* Presentation at the annual convention of the American Psychological Association, Orlando, FL.

Sullivan, A. L., & Long, L. (2010). Examining the changing landscape of school psychology practice: A survey of school-based practitioners' training and involvement in RTI. *Psychology in the Schools, 47,* 1059-1079.

Tichá, R., Espin, C. A., & Wayman, M. M. (2007). 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,* 132-142. doi:10.1111/j.1540-5826.2009.00287.x

Tindal, G. (2013). Curriculum-based measurement: A brief history of nearly everything from the 1970s to the present. *ISRN Education (International Scholarly Research Network), 2013,* 1-29. doi:10.1155/2013/958530

Torgesen, J. K. (2009). The response to intervention instructional model: Some outcomes from a large-scale implementation in Reading First schools. *Child Development Perspectives, 3,* 38-40. doi:10.1111/j.1750-8606.2009.00073.x

U.S. Census Bureau. (2010). *Language use in the United States: 2007.* Retrieved from http://www.census.gov.ezproxy1.lib.asu.edu/prod/2010pubs/acs-12.pdf

U.S. Department of Education. (2003). *Status and trends in the education of Hispanics.* Washington, DC: National Center for Education Statistics.

U.S. Department of Education. (2006). *Assistance to States for the Education of Children with Disabilities and Preschool Grants for Children with Disabilities, 71 Fed. Reg. 156 (August 14, 2006) (to be codified at 34 C.F.R § 300).* Retrieved from http://www.ed.gov/legislation/FedRegister/finrule/2006-3/081406a.html

U.S. Department of Education. (2011). *U.S. Departments of Education and Justice reach settlement with Arizona Department of Education to ensure that potential ELL students are properly identified*. Retrieved from http://www.ed.gov/news/press-releases/us-departments-education-and-justice-reach-settlement-arizona-department-educati

U.S. Department of Education, National Center for Education Statistics. (2012). *The condition of education 2012*. Alexandria, VA: Author.

U.S. Department of Education, National Center for Education Statistics. (2013). *School district demographics system*. Retrieved from http://nces.ed.gov/surveys/sdds/

U.S. Office of Education. (1976). Proposed rule-making. *Federal Register, 41*(230), 52404–52407. Washington, DC: U.S. Government Printing Office.

U.S. Office of Education. (1977). Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. *Federal Register, 42*(250), 65082–65085. Washington, DC: U.S. Government Printing Office.

U.S. Office of Special Education Programs. (2000). *History: Twenty-Five years of progress in educating children with disabilities through IDEA*. Washington, DC: Government Printing Office.

VanDerHeyden, A. M., & Witt, J. C. (2005). Quantifying context in assessment: Capturing the effect of base rates on teacher referral and a problem-solving model of identification. *School Psychology Review, 34*, 161-183.

VanDerHeyden, A. M., & Witt, J. C. (2008). Best practices in can't do/won't do assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., Vol. 2, pp. 131–140). Bethesda, MD: National Association of School Psychologists.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of an RTI model on identification of children for special education. *Journal of School Psychology, 45*, 225-256.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). The development and validation of a process for screening and referrals to special education. *School Psychology Review, 32*, 204-227.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice, 18*, 137-146. doi:10.1111/1540-5826.00070

Vaughn, S., Cirino, P.T., Wanzek, J., Wexler, J., Fletcher, J.M., Denton, C. D., et al. (2010). Response to intervention for middle school students with reading difficulties: effects of a primary and secondary intervention. *School Psychology Review, 39*, 3-21.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to intervention as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69*, 391-409.

Vaughn, S., Linan-Thompson, S., Mathes, P. G., Cirino, P. T., Carlson, C. D., Pollard-Durodola, S. D., . . . Francis, D. J. (2006). Effectiveness of Spanish intervention for first-grade English language learners at-risk for reading difficulties. *Journal of Learning Disabilities, 39*, 56-73. doi:10.1177/00222194060390010601

Vaughn, S., Mathes, P. G., Linan-Thompson, S., Cirino, P., Carlson, C. D., Pollard-Durodola, S., . . . Francis, D. J. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal, 107*, 153-180. doi:10.1086/510653

Vaughn, S., Wexler, J., Leroux, A. J., Roberts, G., Denton, C. A., Barth, A. E., & Fletcher, J. M. (2012). Effects of intensive reading intervention for eighth grade students with persistently inadequate response to intervention. *Journal of Learning Disabilities, 45*, 515-525. doi:10.1177/0022219411402692

Vellutino, F., Scanlon, D., & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. *Journal of Learning Disabilities, 33*, 223-238.

Wagner, R. K., Francis, D. J., & Morris, R. D. (2005). Identifying English language learners with learning disabilities: Key challenges and possible approaches. *Learning Disabilities Research & Practice, 20*, 6-15. doi:10.1111/j.1540-5826.2005.00115.x

Wang, J., & Wang, H. (2007). *Should grade retention decisions be indicators-based or model-driven?* Los Angeles, CA: University of California National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Watson, S. M. R., Rinaldi, C., Navarrete, L., Bianco, M., & Samson, J. (2007). *Infosheet about diverse learners*. Retrieved from http://www.cldinternational.org/InfoSheets/Diverse.asp

Wayman, M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 41*, 85-120.

Weddig, R. A. (1984). Parental interpretation of psychoeducational reports. *Psychology in the Schools, 21*, 477-481.

What Works Clearinghouse. (2014). *What Works Clearinghouse procedures and standards handbook version 3.0.* Retrieved from http://ies.ed.gov/ncee/wwc/

Wiener, J. (1987). Factors affecting educators' comprehension of psychological reports. *Psychology in the Schools, 24*, 116-126. doi:10.1002/1520-6807(198704)24:2%3C116::AID-PITS2310240205%3E3.0.CO;2-2

Wilkinson, C. Y., Ortiz, A. A., Robertson, P. M., & Kushner, M. I. (2006). English language learners with reading-related LD: Linking data from multiple sources to make eligibility determinations. *Journal of Learning Disabilities, 39*, 129-141. doi:10.1177/00222194060390020201

Wisconsin Department of Education. (2011). *WI-RtI brief: Specific learning disabilities rule implementation and RtI*. Retrieved from http://www.wisconsinrticenter.org/

Witt, J., & Beck, R. (1999). *One minute academic functional assessment and interventions: "Can't" do it…or "won't" do it?"* Longmont, CO: Sopris West.

Witt, J. C., Daly, E., & Noell, G. (2000). *Functional Assessments*. Longmont, CO: Sopris West.

Wixson, K. (2011). A systematic review of RTI research. *The Elementary School Journal, 111*, 503-510.

Wodrich, D. L., & Schmitt, A. J. (2006). *Patterns of learning disorders: Working systematically from assessment to intervention.* New York, NY: Guilford Press.

Wolf, M. K., Kao, J. C., Griffin, N., Herman, J. L., Bachman, P. L., Chang, S. M., & Farnsworth, T. (2008). *Issues in assessing English language learners: Englishlanguage proficiency measures and accommodation uses—practice review*. Los Angeles, CA: University of California National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Woodcock, R. W., & Muñoz-Sandoval, F. (2011). *Woodcock-Muñoz Language Survey Revised - Normative Update*. Itasca, IL: Riverside Publishing.

Xu, Y., & Drame, E. R. (2008). Culturally appropriate context: Unlocking the potential of response to intervention for English language learners. *Early Childhood Education, 35*, 305-311. doi:10.1007/s10643-007-0213-4

Yell, M. L., & Drasgow, E. (2007). Assessment for eligibility under IDEIA and the 2006 regulations. *Assessment for Effective Intervention, 32*, 202-213. doi:10.1177/15345084070320040201

Yell, M. L., Shriner, J. G., & Katsiyannis, A. (2006). Individuals with Disabilities Education Improvement Act of 2004 and IDEA regulations of 2006: Implications for educators, administrators, and teacher trainers. *Focus on Exceptional Children, 39*, 1-24.

Yell, M. L., & Walker, D. W. (2010). The legal basis of response to intervention: Analysis and implications. *Exceptionality, 18*, 124-137.

Zehler, A., Fleischman, H., Hopstock, P., Stephenson, T., Pendzick, M., & Sapru, S. (2003). *Descriptive study of services to LEP students and LEP students with disabilities* (Vol. 1). Arlington, VA: Development Associates Inc.

Zirkel, P. A. (2013). Recent legal developments of interest to special educators. *Intervention in School and Clinic, 48*, 319-322. doi:10.1177/1053451212472234

Zirkel, P. A., & Thomas, L. (2010a). State laws and guidelines for implementing RTI. *Teaching Exceptional Children, 43*, 60-73.

Zirkel, P. A., & Thomas, L. (2010b). State laws for RTI: An updated snapshot. *Teaching Exceptional Children, 42*, 56-63.

APPENDIX A

CODING PROTOCOL

		Item	Report Number				
			CLD☐	CLD☐	CLD☐	CLD☐	CLD☐
Basic Information	1.	Gender (M = Male, F = Female)	☐M ☐F	☐M ☐F	☐M ☐F	☐M ☐F	☐M ☐F
	2.	Age at time of report					
	3.	Grade					
	4.	Race/Ethnicity (W = White, A = Asian, B = Black, H = Hispanic, I = American Indian, O = Other)	☐W ☐H ☐A ☐I ☐B ☐O	☐W ☐H ☐A ☐I ☐B ☐O	☐W ☐H ☐A ☐I ☐B ☐O	☐W ☐H ☐A ☐I ☐B ☐O	☐W ☐H ☐A ☐I ☐B ☐O
	5.	School year in which the report was completed	☐2009-2010 ☐2010-2011 ☐2011-2012 ☐2012-2013	☐2009-2010 ☐2010-2011 ☐2011-2012 ☐2012-2013	☐2009-2010 ☐2010-2011 ☐2011-2012 ☐2012-2013	☐2009-2010 ☐2010-2011 ☐2011-2012 ☐2012-2013	☐2009-2010 ☐2010-2011 ☐2011-2012 ☐2012-2013
	6.	Degrees/licensure status of psychologist	☐MA ☐PhD ☐EdD ☐PsyD ☐Licensed ☐Other: ____	☐MA ☐PhD ☐EdD ☐PsyD ☐Licensed ☐Other: ____	☐MA ☐PhD ☐EdD ☐PsyD ☐Licensed ☐Other: ____	☐MA ☐PhD ☐EdD ☐PsyD ☐Licensed ☐Other: ____	☐MA ☐PhD ☐EdD ☐PsyD ☐Licensed ☐Other: ____
	7.	Was the psychologist bilingual?	☐Y ☐N	☐Y ☐N	☐Y ☐N	☐Y ☐N	☐Y ☐N
	8.	Cognitive instruments utilized *- Include instrument even if only some subtests of the measure were given; exception is KABC-2, which has a separate listing if only NVI was given* *- Specify language of administration if not English*	☐WISC-IV ☐WJ-III Cog ☐KABC-2 ☐KABC-2 NVI Only ☐SB-5 ☐DAS-2 ☐UNIT ☐Leiter ☐CTONI ☐Other: ____	☐WISC-IV ☐WJ-III Cog ☐KABC-2 ☐KABC-2 NVI Only ☐SB-5 ☐DAS-2 ☐UNIT ☐Leiter ☐CTONI ☐Other: ____	☐WISC-IV ☐WJ-III Cog ☐KABC-2 ☐KABC-2 NVI Only ☐SB-5 ☐DAS-2 ☐UNIT ☐Leiter ☐CTONI ☐Other: ____	☐WISC-IV ☐WJ-III Cog ☐KABC-2 ☐KABC-2 NVI Only ☐SB-5 ☐DAS-2 ☐UNIT ☐Leiter ☐CTONI ☐Other: ____	☐WISC-IV ☐WJ-III Cog ☐KABC-2 ☐KABC-2 NVI Only ☐SB-5 ☐DAS-2 ☐UNIT ☐Leiter ☐CTONI ☐Other: ____
	9.	Academic instruments utilized *- Include instrument even if only some subtests of the measure were given* *- Specify language of administration if not English*	☐WJ-III ☐WIAT-III ☐TERA-3 ☐TEMA-3 ☐TEWL-3 ☐YCAT ☐Other: ____	☐WJ-III ☐WIAT-III ☐TERA-3 ☐TEMA-3 ☐TEWL-3 ☐YCAT ☐Other: ____	☐WJ-III ☐WIAT-III ☐TERA-3 ☐TEMA-3 ☐TEWL-3 ☐YCAT ☐Other: ____	☐WJ-III ☐WIAT-III ☐TERA-3 ☐TEMA-3 ☐TEWL-3 ☐YCAT ☐Other: ____	☐WJ-III ☐WIAT-III ☐TERA-3 ☐TEMA-3 ☐TEWL-3 ☐YCAT ☐Other: ____

	10.	Social-emotional instruments utilized - *Specify language of administration if not English*	☐BASC-2 ☐Conners 3 ☐CAB ☐Other: ____		☐BASC-2 ☐Conners 3 ☐CAB ☐Other: ____		☐BASC-2 ☐Conners 3 ☐CAB ☐Other: ____		☐BASC-2 ☐Conners 3 ☐CAB ☐Other: ____		☐BASC-2 ☐Conners 3 ☐CAB ☐Other: ____	
	11.	Other instruments utilized - *Include proficiency tests given for the purposes of the current evaluation* - *Specify language of administration if not English*	☐Vineland ☐BVAT ☐VMI ☐Other: ____		☐Vineland ☐BVAT ☐VMI ☐Other: ____		☐Vineland ☐BVAT ☐VMI ☐Other: ____		☐Vineland ☐BVAT ☐VMI ☐Other: ____		☐Vineland ☐BVAT ☐VMI ☐Other: ____	
Review of Existing Data	10.	Does the evaluation include the following? (ARS §300.305)										
		-Review of existing evaluations (or a statement that there were no previous evaluations)	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Review of other information provided by parents	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Review of current classroom-based performance	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Review of local and state-wide assessments (or a statement that this data wasn't available)	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Classroom-based observations by the examiner	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Observations from teacher/other service providers (can be in the form of teacher comments re: the student)	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Relevant functional information (ARS §300.304) (can include academic/behavioral)	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Relevant developmental information (ARS §300.304)	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
Overall Disability Rule Outs	11.	Does the evaluation address the following rule outs? (ARS §300.306 & ARS §300.307)										
		(a) Lack of appropriate instruction in reading	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is this primarily a disclaimer?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is supporting evidence provided to demonstrate that prior to, or as part of, the referral process, the child was provided appropriate instruction in regular education settings, delivered by qualified personnel (e.g., information on curriculum, specific interventions)?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is there data-based documentation of repeated assessments of achievement at reasonable intervals, reflecting formal assessment of student progress during instruction (e.g., class benchmark scores, RTI data)?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		(b) Lack of appropriate instruction in math	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is this primarily a disclaimer?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N

		-Is supporting evidence provided to demonstrate that prior to, or as part of, the referral process, the child was provided appropriate instruction in regular education settings, delivered by qualified personnel (e.g., information on curriculum, specific interventions)?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is there data-based documentation of repeated assessments of achievement at reasonable intervals, reflecting formal assessment of student progress during instruction (e.g., class benchmark scores, RTI data)?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		(c) Limited English proficiency	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is this primarily a disclaimer?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is the student's language proficiency described? (addresses level of proficiency rather than simply stating is or is not proficient)	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
Assessment Tools	12.	Were a variety of assessment tools and strategies used to gather relevant functional, developmental, and academic information about the child? (ARS §300.304)	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
	13.	Does the report state that evaluation procedures were selected and administered so as not to be discriminatory on a racial or cultural basis? (ARS §300.304)	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is this primarily a disclaimer?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is the individual's racial/cultural background described (not simply stated; may address acculturation, etc.)?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is the rationale for the selection of procedures provided (e.g., test was normed on representative sample)?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
Statement of SLD and SLD Rule Out	14.	Does the report include the following? (ARS §300.311)										
		-Whether the child has SLD	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is this based on severe discrepancy? (ARS §300.307)	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is this based on RTI? (ARS §300.307)	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-The basis for making the determination, including that it was made in accordance with IDEA/state law	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Any relevant behaviors noted during the observation and the relationships of the behavior to academic functioning	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Any educationally relevant medical findings	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Whether the child does not achieve adequately for his/her age or meet state-approved grade level standards	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Whether the child does not make progress sufficient to meet age or state-approved grade level standards	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-The child's patterns of strengths and weaknesses	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N

	15.	Does the evaluation address the following SLD rule outs? (ARS §300.309; in addition to previous disability rule outs)																			
		(a) Visual, hearing, or motor disability (sensory deficits are considered visual/hearing)	V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N		
		-Is this primarily a disclaimer?	V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N		
		-Are these factors discussed? (e.g., vision/hearing or motor test results)	V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N			V H M	☐Y ☐N ☐Y ☐N ☐Y ☐N		
		(b) Mental retardation/intellectual disability	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		-Is this primarily a disclaimer?	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		-Are intellectual factors discussed?	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		(c) Emotional disturbance	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		-Is this primarily a disclaimer?	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		-Are emotional factors discussed?	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		(d) Cultural factors	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		-Is this primarily a disclaimer?	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		-Are cultural factors discussed?	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		(e) Environmental or economic disadvantage	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		-Is this primarily a disclaimer?	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		-Are environmental/economic factors discussed?	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
	16.	What SLD eligibility categories did the child fall into? (ARS §300.309) (BR = Basic Reading, RF = Reading Fluency, RC = Reading Comprehension, MC = Math Calculation, MP = Math Problem Solving, WE = Written Expression, OE = Oral Expression, LC = Listening Comprehension)	☐BR ☐RF ☐RC ☐MC	☐MP ☐WE ☐OE ☐LC			☐BR ☐RF ☐RC ☐MC	☐MP ☐WE ☐OE ☐LC			☐BR ☐RF ☐RC ☐MC	☐MP ☐WE ☐OE ☐LC			☐BR ☐RF ☐RC ☐MC	☐MP ☐WE ☐OE ☐LC			☐BR ☐RF ☐RC ☐MC	☐MP ☐WE ☐OE ☐LC		
RTI	17.	Is there documentation of methods, strategies, and interventions that have been tried and were not successful?	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		-Is this primarily a disclaimer?	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		-Are specific accommodations and/or strategies listed?	☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N			☐Y	☐N		
		-Specific accommodations/strategies utilized (e.g., preferential seating)																			

		-Are specific interventions listed?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Specific interventions utilized (e.g., Wilson reading). ADE (2012) defines an intervention as "the directing of instruction in the area(s) of concern that is in addition to regular classroom instruction. Interventions are designed to meet the identified needs of an individual and are monitored on regular and frequent basis."										
	18.	Is universal screening data included?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-What type of universal screening data is presented? (ORF WPM = oral reading fluency words per minute; Mazes = number of correct responses on a maze task; MC Digits = math calculation digits correct)	☐ORF WPM ☐Mazes ☐MC Digits ☐Other: ____		☐ORF WPM ☐Mazes ☐MC Digits ☐Other: ____		☐ORF WPM ☐Mazes ☐MC Digits ☐Other: ____		☐ORF WPM ☐Mazes ☐MC Digits ☐Other: ____		☐ORF WPM ☐Mazes ☐MC Digits ☐Other: ____	
		-How many data points are provided?	ORF WPM: __ Mazes: __ MC Digits: __ Other: __		ORF WPM: __ Mazes: __ MC Digits: __ Other: __		ORF WPM: __ Mazes: __ MC Digits: __ Other: __		ORF WPM: __ Mazes: __ MC Digits: __ Other: __		ORF WPM: __ Mazes: __ MC Digits: __ Other: __	
	19.	Is there evidence that a can't do/won't do assessment was administered?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
	20.	Is progress monitoring data from targeted and/or intensive interventions included?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-What type of progress monitoring data is presented? (ORF WPM = oral reading fluency words per minute; Mazes = number of correct responses on a maze task; MC Digits = math calculation digits correct)	☐ORF WPM ☐Mazes ☐MC Digits ☐Other: ____		☐ORF WPM ☐Mazes ☐MC Digits ☐Other: ____		☐ORF WPM ☐Mazes ☐MC Digits ☐Other: ____		☐ORF WPM ☐Mazes ☐MC Digits ☐Other: ____		☐ORF WPM ☐Mazes ☐MC Digits ☐Other: ____	
		-Is data presented in the form of one or more graphs?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-How many data points are provided?	ORF WPM: __ Mazes: __ MC Digits: __ Other: __		ORF WPM: __ Mazes: __ MC Digits: __ Other: __		ORF WPM: __ Mazes: __ MC Digits: __ Other: __		ORF WPM: __ Mazes: __ MC Digits: __ Other: __		ORF WPM: __ Mazes: __ MC Digits: __ Other: __	
	21.	Does the report include information on the length (e.g., 12 weeks), frequency (e.g., three days/week), and timing (e.g., 30 minutes/session) of the interventions?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-What was the length/frequency/timing for each intervention implemented?										

	22.	Does the report state who implemented each intervention (e.g., special education teacher, general education teacher, aide, computer-based)?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Who implemented each intervention?										
	23.	Was a specific decision rule referenced to determine whether or not the student had responded (e.g., the slope of the student's trend line was insufficient; the student performed below the goal line for several weeks)?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		DO NOT COMPLETE THE PORTION BELOW IF HOME/DOMINANT/FIRST LANGUAGES ARE ENGLISH AND NO ELL STATUS INDICATED										
Language	24.	Is the primary language used in the student's home identified? If so, what is it?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
	25.	Is the language most often spoken by the student identified? If so, what is it?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
	26.	Is the student's first language identified? If so, what is it?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
	27.	Does the report state how long the student has lived in the United States?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
	28.	Is the impact of language proficiency on educational progress discussed?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
	29.	Were evaluation procedures provided and administered in the child's native language or other mode of communication and in a form most likely to yield accurate information about what the child knows and can do academically, developmentally and functionally? (ARS §300.304) (can include nonverbal assessment)	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is the most proficient language identified? If so, what is it?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is the process for determining proficiency described?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is the state English proficiency evaluation (AZELLA) discussed?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Was an individually-administered proficiency test (e.g., Woodcock-Muñoz Language Survey) administered as part of this evaluation? If so, which test?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
		-Is additional individually-administered proficiency test data included (e.g., from a speech evaluation)? If so, which test?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N
	30.	Is the student's educational history discussed?	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N	☐Y	☐N

APPENDIX B

IRB APPROVAL LETTER

ASU Knowledge Enterprise Development

EXEMPTION GRANTED

Linda Caterino Kulhavy
Division of Educational Leadership and Innovation - Tempe
480/965-7524
Linda.Caterino@asu.edu

Dear Linda Caterino Kulhavy:

On 11/25/2013 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Integration of Traditional Assessment and Response to Intervention Data in Psychoeducational Evaluations of Culturally and Linguistically Diverse Students
Investigator:	Linda Caterino Kulhavy
IRB ID:	STUDY00000293
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	None

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (1) Educational settings on 11/25/2013.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator