The Report committee for Sarah Michelle Pettit
Certifies that this is the approved version of the following report:

The Effect of Response to Intervention and Discrepancy Testing in Diagnosing English Language Learners with a Reading Learning Disability

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Cindy Carlson

Janay Sander

The Effect of Response to Intervention and Discrepancy Testing in Diagnosing English Language Learners with a Reading Learning Disability

by

Sarah Michelle Pettit, B.A.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May 2011

Abstract

The Effect of Response to Intervention and Discrepancy Testing in Diagnosing English Language Learners with a Reading Learning Disability

by

Sarah Michelle Pettit, M.A.

The University of Texas at Austin, 2010

SUPERVISOR: Cindy Carlson

This study looks at two methods of identification for a specific learning disability in reading – Response to Intervention and Discrepancy Testing – and examines their efficacy with English Language Learners (ELL). A McNemar chi square analysis will be used to compare the outcomes of the two methods. Additionally, a multiway contingency table will be constructed and the association between English as a First Language (EFL) students and ELL students will be determined by using a conditional odds ratio. It is expected that both methods of learning disability testing will have difficulty distinguishing between a learning disability in reading and a language acquisition problem. The results will help prevent the misplacement of ELLs into Special Education.

Table of Contents

Introduction

Although the United States has always been a country of immigrants, the number of immigrants entering the country in recent years has almost reached the peaks of immigration in the 1800s and early 1900s. It is estimated that by the year 2030, about 40% of the school population will speak English as a second language (Klinger et al., 2006). Non-English speaking students are the fastest growing subgroup of children among public school populations, with an annual increase of approximately 10 percent (McCardle et al., 2005). Currently, about 3.4 to 3.9 million students are English Language Learners (ELL) (State Education Agency, 2001; US Census 2000). English Language Learners are defined as students whose English skills are so limited that they cannot profit from general education instruction provided in English without special language support (Wilkinson et al., 2006). In the past, such students have been referred to as Limited English Proficient (LEP), although that term is now out of vogue. Given that Mexico is the largest source country for U.S. immigration, about two-fifths of foreign-born children are of Mexican descent, and three quarters of all ELL students between pre-kindergarten to fifth grade speak Spanish. (Capps et al., 2006).

ELLs are between 1.42 and 2.43 times more likely to be placed in programs LD and speech and language impairments than their English-speaking peers (Liu et al., 2008). Special education law requires a distinction between students who have limited proficiency in English and students who have a learning disability (LD) (Barrera, 2006). However, ELL students and those with LDs appear to share many characteristics.

(Barrera, 2006). For example, students who are second language learners often exhibit severe discrepancies between their academic potential and their actual achievement (Barrera, 2006). It is important to be able to accurately identify the ELL/LD population in order to know how to best serve them (McCardle et al., 2005).

Currently, most cognitive tests are not appropriate for ELLs students, and intelligence tests may underestimate the potential of students whose primary language is not English (Klinger et al. 2006, Liu et al., 2008). The norming standards for these tests typically exclude children from ELL backgrounds (Lopez et al., 1997). Most students are tested in English regardless of their home language, and without accommodations (Klinger et al., 2006). As a result, they typically yield scores that reflect English language knowledge instead of cognitive ability (Lopez et al. 1997). Only 1% of school psychologists surveyed attempted to determine if a discrepancy occurred in both English and the student's home language (Klinger et al. 2006). However, language loss in the primary language is common when LEP children focus attention on learning a second language (Lopez et al., 1997), so testing in both languages may still not provide an accurate portrait of a possible discrepancy.

Response to Intervention offers a promising alternative for reducing the disproportionate number of culturally and linguistically diverse students in special education by identifying students at risk early and providing preventative instruction to accelerate progress (Linan-Thompson et al, 2007). However, research based interventions tailored to ELL students are still rare. A study by Linan-Thompson et al. (2007) found that among first grade ELLs who receive RTI, 80 percent did not meet

benchmark criteria after a year. RTI measures achievement with curriculum-based measurements (Barrera, 2006). This approach examines a student's learning ability as a function of what the student can do as he or she is being taught, as opposed to what the student does or does not know.

This study will examine whether there is a difference in the proportion of ELL students that RTI or discrepancy testing identify as having a reading LD. Are the two tests measuring the same construct in ELL students? The study will then see degree of association between how the proportion of ELL students identified by these two methods compare to the proportion of English as a First Language (or EFL) students identified by these methods. The results will show how generalizable these different methods of reading LD identification are with an ELL population.

Integrative Analysis

This integrative analysis will examine conceptualizations learning disabilities (LDs) and the current methods by which they are identified. This analysis will look at the discrepancy method of identification and the Response to Intervention (RTI) method as two possible ways for diagnosing an LD. Further more, the analysis will look at various characteristics of English Language Learners (ELLs), particular factors that may undermine their achievement in school. Finally, this integrative analysis will discuss the efficacy of the two methods of LD identification with the ELL population.

Learning Disabilities (LD)

Definition of LDs

Given that most students who receive special education services in school are diagnosed with a learning disability (LD), one might think that there would be a clear and precise definition describing the construct. After all, just over 50% of all students in the United States that are identified for special education in the are classified as LD which is approximately 5% of the school-age population (Vaughn & Fuchs, 2003). However, there is currently no consensus in the field on how to define them. Flanagan and Ortiz (2006) compare the situation to building the Tower of Babel, saying:

> The task of delineating an LD definition and a corresponding set of diagnostic criteria suffers from the very same problem—failure to use a common language and an inability to communicate within and across disciplines using a common set of terms that mean the same thing to all who use them.

However, what stands out from most definitions is that students identified as LD "have learning problems that cannot be attributed to lack of intelligence" (Shepard, 2001). Virtually all LD definitions rest on this discrepancy between achievement and ability. Part of the problem is that many scientific and legal definitions define what an LD *is not*, but not what it *is*. For example, most scientific and legal prescriptions for LD identification require an evaluation of exclusionary factors such as cultural differences, linguistic differences, economic disadvantage, emotional or psychological disturbance, lack of motivation, poor instruction, etc. If the academic or cognitive deficits can be ascribed to any of these influences, then an LD diagnosis is inappropriate. This vague definition is problematic in research because diagnostic criteria vary so widely that the LD samples are very heterogeneous. This puts the generalizability of results into question.

However, in spite of the fact that there is no formally recognized definition for an LD, a review of the literature (Flanagan et al., 2006) shows that most researchers do share common components in defining an LD. These components include a history of academic difficulty, use of pre-referral interventions, identified academic deficits, identified cognitive deficits, intact cognitive abilities in areas not strongly related to the academic deficits, underachievement, an evaluation of exclusionary factors, and evidence of functional impairment.

Additionally, there is considerable interstate variability in terms of prevalence of students diagnosed with an LD, though there are consistent diagnoses from year to year within states. That indicates that while states are interpreting the definition of an LD

differently, they are doing so consistently. A study (Hallahan, Keller, Martinez, Byrd, Gelman & Fan, 2007) compared interstate variability of prevalence rates for special education categories from 1984-85 through 2001-02. The study found that LDs, which were thought to be the most variable special education category, were in fact consistently the *least* variable. The data suggests that high-incidence categories, such as LDs, may be just as well defined, if not better defined, than low-incidence categories. The downward trend in variability fits with the notion that policy makers and practitioners become more consistent the longer they engage in defining and identifying students with a particular disability. However, for example, three times as many students are identified as learning disabled in Rhode Island than in Kentucky. While it may be argued that there might still be considerable state-to-state variability even with perfect identification procedures, clearly the vague criteria used to describe an LD is, in part, to blame.

Discrepancy Method of Identification

The most traditional method of deciding whether or not a student has an LD is called the "discrepancy method of identification." This method is based on multiple regression and uses the expected performance predicted from aptitude scores so that children from the full ability continuum will be identified (Shepard, 2001). It involves giving both an ability and achievement test, which are normed together. An anticipated achievement score is computed for each child based on ability, grade level, and sex. Then, for each ability score, the 10% whose actual achievement is most discrepant from their anticipated achievement are identified as likely LD; therefore, LD children are clearly distinguished from slow learners. The discrepancy method of identification is

based on the theory that an LD exists if a child's academic achievement lags significantly behind intellectual ability and there is no other known cause for the discrepancy.

There have been some technical advances with traditional discrepancy models. Discrepancy now provides a far more dynamic view of cognitive ability than previous conceptions, which were based on a singular global ability. A student may also have a significant discrepancy between a global score from an IQ test and an achievement score from an academic skills test. The problem with this type of interpretation is that a global IQ scores is a good predictor of general achievement, and is frequently attenuated by the abilities that comprise it, which lead to low achievement.

The most advanced use of the discrepancy method of identification is based in Cattell-Horn-Carroll (CHC) theory, and is called Cross-Battery Assessment approach. The CHC model features ten broad abilities and over seventy narrow abilities, with each broad ability subsuming two or more narrow abilities. Cross-battery assessment employs different test batteries and which one must then interpret to provide a comprehensive assessment. This theory specifies the relations between cognitive abilities and academic abilities (Flanagan et al., 2006). For example, reading involves numerous cognitive abilities, so identification of the specific abilities that are impeding the ability to read would have significant implications for instructional planning and interventions.

So what does the cognitive and academic profile look like for a student who has an LD? A student could have relatively normal functioning in most cognitive areas and a deficit in an area that is strongly related to an area of academic deficiency. Flanagan et al. (2006) describe this as "below average cognitive and academic consistency within an

otherwise normal ability profile." When a student does not meet the discrepancy criteria, there is a strong possibility that the student is a "slow learner" (SL; i.e., a student with an IQ level between about 70 and 85). About 14% of the school population may be deemed SL, but this group does not demonstrate unexpected learning failure because their achievement level is consistent with IQ level. However, while it is conceptually and methodologically superior to other approaches, the discrepancy model is thought by many to be deficient as a sole criterion for LD identification.

There are a few weaknesses associated with the discrepancy model of identifying LDs. Misidentification of typical children as LD will occur both because of statistical artifacts (Fletcher, Denton, & Francis, 2005) and because there are other real causes of discrepant profiles. For example, prolonged absence from school or a nontraditional curriculum that greatly underemphasizes basic skills could create a profile that mimics an LD. There is also measurement error in both cognitive and achievement instruments. While some argue that further clinical diagnosis will sort out the true LD cases from the rest, according to Shepard (2001), there is little valid evidence to demonstrate that experts can tell which is which within this pool of cases (students who score a discrepancy but do not have an LD).

Also, assumptions underlying the IQ-achievement discrepancy model have not been supported. It has not been empirically supported that the discrepancy from the IQ would meaningfully relate to the severity of the LD, that academic performance of students with a discrepancy differs from that of students without a discrepancy, that a discrepancy yields reliable information, that the findings inform instruction, or that the

use of IQ tests is a necessary procedure for identifying students with an LD. Two meta-analyses have found negligible differences between a discrepant full-scale IQ and reading achievement group versus poor readers without an IQ-achievement discrepancy (Hoskyn & Swanson, 2000; Stuebing et al., 2002). While some argue that cognitive information is useful, "there is little evidence that instruction addressing strengths and weaknesses in cognitive skills is related to intervention outcomes" (Fletcher, et al, 2003). The presence or absence of an IQ-achievement discrepancy does not lead to any differences in long-term prognosis when compared to low achievement (Francis, Shaywitz, Stuebing, Shaywitz & Fletcher, 1996). With time, "individuals with increasingly severe academic problems will show increasingly flat profiles on cognitive tasks (and achievement measures) in direct correspondence to severity."

A major issue with the discrepancy method is that a student must demonstrate extreme difficulty in learning before teachers recognize the problem and refer the student for special education. This leads to late identification for students with special needs, imprecise screening, false negatives (i.e., unidentified students) who are either not provided necessary services or provided services too late, and use of identification measures that are not linked to instruction (Vaughn & Fuchs, 2003). Several researchers have found that LD children have significantly below-average IQ scores, even though IQ by definition should be uncorrelated with LD.

Fletcher et al (2005) did a study reviewing the validity of models based on aptitude-achievement discrepancies, low achievement, intra-individual differences, and response to instruction for the classification and identification of LDs. They found that

models based on aptitude-achievement discrepancies and intra-individual differences showed little discriminate validity. Low achievement models had higher discriminant validity, but they did not adequately address the unexpected underachievement component of the LD construct. Response to instruction (which will be discussed in greater detail below) has higher reliability and validity because it does not rely on one measurement at a single time point. Hybrid models (Fletcher et al, 2005) that combined low achievement with response to instruction both capture the LD construct and relate to instruction. The study assumed that the construct of an LD is a latent variable that is imperfectly measured by the different approaches.

RTI Method of Identification

The roots of response-to-intervention method (RTI) of the identification of LD come from a 1982 National Research Council Study that determined that validity of special education classification should be judged according to three criteria: (1) whether the quality of general education is such that adequate learning might be expected; (2) whether the special education program is of sufficient value to improve student outcomes and thereby justify the classification; and (3) whether the assessment process used for identification is accurate and meaningful. The core concepts of RTI include: (a) application of scientific, research-based interventions in general education settings; (b) measurement of a student's response to those interventions; and (c) the use of the data to inform instruction. RTI was developed to have three phases. The first phase consists of high-quality, research-based instruction for all students in the general education classroom. The students who do not sufficiently advance with "generally effective"

instruction receive small-group, high-intensity intervention provided either within the general education environment or outside the classroom through a "special education-like" instructional model. Those who fail to respond to this intervention enter the third tier, which prompts eligibility for placement into special education.

Measuring progress relies on curriculum-based measurements. Risk is defined as dual discrepancy between curriculum-based measurement (CBM) level and CBM rate of growth between the target student and his or her classmates. In phase three, CBM is used to measure a student's responsiveness to classroom adaptations. If the child demonstrates responsiveness, then the presence of a disability has been disconfirmed, and the general education program, with or without adaptations, continues so that the student's needs are addressed. If the child fails to respond, then a more highly differentiated and intensive instructional program—special education—is required.

There are several benefits to using CBM. CBM represents an assessment method that can provide the multiple sources of documentation needed for modeling academic growth, distinguishing between ineffective general education environments and unacceptable individual student learning, informing instructional planning, and evaluating relative instruction effectiveness. Research has reported positive outcomes stemming from the identification procedures five years after district-wide implementation. Students received special education services at younger ages, and disproportionality was reduced in new special education placements.

Currently, there are two different models of RTI: the Problem Solving Model (PSM) and the Standard Treatment Protocol (STP). PSM is a "systematic, data-driven

process that is designed to use collaborative teaming […] emphasizing early classroom interventions, goal setting, data-based decision making, and functional evaluation procedures" (Hollenbeck, 2007). STP standardizes the intervention for all struggling learners (Hollenbeck, 2007). It is used most frequently in reading research. A third model, a mixture of PSM and STP dubbed the Mixed Model is sometimes used as well. One study (Hollenbeck, 2007) with a Mixed Model found that special education referrals between 1996 and 2004 across 36 schools dropped 41% in kindergarten, 34% in first grade, 25% in second, and 19% in third. It is not known if there is a rebound effect in higher grades.

The RTI system has several strengths. For example, it ensures pre-referral interventions, which many practitioners of LD identification do not systematically employ (Flanagan et al., 2006). Pre-referrals are an important part of the LD identification process as they can eliminate false positives, suspected cases that are in fact due to lack of appropriate instruction, poor motivation, or other factors. RTI could also generate specific data regarding a student's learning difficulties, allowing for the creation of specialized Individualized Educational Programs (IEPs) for those who enter special education (Hollenbeck, 2007). Research showed that students who received early intervention in kindergarten showed moderately high group differences at the end of third grade. RTI models also eliminate teacher bias in referral, impacting the disproportionality of minority students and males in special education (Hollenbeck, 2007).

In addition, there is evidence that inadequate responders to early reading intervention differed from responders in both pre-intervention and achievement scores and pre-intervention cognitive tasks. Poor responders "showed brain activation patterns that generally demonstrated a failure to activate left hemisphere areas known to be involved in the development of reading skills. In fact, those students who were inadequate responders to reading intervention showed predominant right hemisphere activity much like that observed in children and adults with identified reading disabilities" (Fletcher, Denton & Francis, 2003).

There are two schools of thought towards the role of RTI in determination of an LD: data could be used as a component of a comprehensive evaluation or as a replacement for the traditional IQ-achievement discrepancy. However, it is unclear whether the main goal of RTI is specific LD identification or providing effective instruction. Researchers (Kavale et al, 2007) have several concerns about using RTI as a method of identifying LDs. One problem is that research in RTI is limited since its implementation has been fairly recent. Specifically, there is some question around "instruction" that involves relatively intensive (although short-term) tutoring using a standard protocol. If a student responds to this relatively intensive instruction, has the presence of a disability (and a need for special education) been disconfirmed? In one study (Hollenbeck, 2007), only 9 of 14 students continued to excel in the general education classroom after meeting exit criteria at the end of 20 weeks of small-group daily reading instruction.

What, then, is the optimal length of supplemental instruction before a student is judged eligible for special education? What options are available for students who fail to thrive after leaving the supplemental program? And what criteria should be used to determine whether or not a student is responding? In one study (Hollenbeck, 2007), one measure identified 2% of a sample to be non-responders, whereas a different measure identified 32% of the sample to be non-responders. Additionally, there are concerns with false positives, or students that are identified as learning disabled who are in fact not. Kavale et al (2008) say "a prevention model increases false positives—always, inevitably, with mathematical certainty." In that vein, there has been some evidence that kindergarten-screening devices tend to over-identify.

Another issue some (e.g., Hallahan, et al., 2007) have with RTI is that it alters the definition of the LD construct. They see RTI as redefining LDs as low achievement, as opposed to low achievement *relative to ability*. Unlike discrepancy, which validates the presence or absence of an accepted construct (i.e. underachievement), RTI can only validate the self-evident fact that a student is experiencing a reading problem. Specifically, use of RTI alone makes it difficult to: (a) distinguish LD from mild mental retardation; (b) distinguish LD students from slow learners; (c) identify intra-individual differences; (d) determine the meaning of positive RTI; and (e) identify the best means to implement effective interventions (Kavale, Kauffman, Bachmeier, & LeFever, 2008).

Another concern is variability in LD identification when using RTI. Hallahan, et al (2007) foresees that some localities or states will elect to use one or the other approach (PSM or STP). Additionally, there may be more variability among states using a PSM

because it relies more on clinical judgment than does the more formulaic STP. At the moment, it appears that PSM is the more widely used of the two.

Combined RTI and Discrepancy

While the debate regarding the use of RTI to diagnose a learning disability as opposed to using norm-referenced testing rages on, not all believe that the two approaches are mutually exclusive (Flanagan, et al., 2006). Even critics of RTI have suggested "RTI should be considered to be one important element within the larger context of the LD determination process," although "RTI as one component of LD determination is insufficient as a sole criterion for accurately determining LD"(Kavale et al, 2007). However, when using the hybrid model, "students who are so classified appear to differ from students with other forms of low achievement." Kavale et al (2007) recommend that researchers and practitioners: (1) reform RTI into what it is, i.e., pre-referral intervention; (2) involve special education only after RTI failure, when the process shifts from prevention to identification; (3) base identification on comprehensive psychometric assessment designed to provide both diagnostic and instructional data.

English Language Learners

Demographics of ELLs

More immigrants emigrate to the United States than to any other nation in the world, with 1.2 million individuals entering the country in 2000, and 20 million arriving in the United States since the passage of the 1965 Hart-Cellar Act (1965), which reduced restrictions on non-European immigration (Turney & Kao, 2009). At the moment, 16% of all children under age 10 are born to immigrant parents in the United States (Palacios,

Guttmannova, & Chase-Landsdale, 2008). The number of students that are English language learners (ELLs) is predicted to grow to 40% of the school-age population by 2030 (Guglielmi, 2008). Although ELL students speak over 460 languages, Spanish is the native language for more than 79% of them (Guglielmi, 2008).

Achievement Trajectories of ELL Students

One of the characteristics that ELL students share with LD students is that both groups tend to have lower achievement in general education. However, research on the achievement trajectories of ELL students are mixed. In other words, not all ELL students have lower achievement in the classroom. ELL students are affected by numerous factors, including how many generations they have been in the United States and a variety of other background variables. While some immigrant students do very well, 66% of the ELL population in the United States scores below the basic reading level in fourth grade and 67% in eighth grade (Cardenas-Hagen, Carlson, & Pollard-Durodola, 2007). This high rate of low achievement may be a reason that ELLs receive special education services more frequency than do EFL children. Most of these referrals are for academic problems, such as difficulties with reading (Cardenas-Hagen et al., 2007).

This phenomenon of lower achievement (as measured by standardized test scores) for ELL students occurs across different races and ethnicities (Leventhal, Xue, & Brooks-Gunn, 2006). One study (Leventhal et al., 2006) examined longitudinal differences in children's verbal abilities as a function of immigrant and racial/ethnic status and found that although all children's mean verbal scores increased with age, immigrant children (except for Black Americans) had lower scores than respective nonimmigrant children.

Mexican-American nonimmigrants had higher verbal scores than their immigrant peers, with only family characteristics explaining a substantial portion of this difference. In another study (Palacios et al., 2008) that looked at student achievement by the number of generations they had been in the United States, Asian students outperformed all other race/ethnic groups in math and reading achievement. The achievement of Black and Hispanic students lagged behind that of White students at the start of kindergarten, with the achievement of Black students continuing to decrease as students matured. However, the early advantage experienced by Asian children also decreased (Palacios et al., 2008; (Han, 2008).

What leads to these differences in achievement? Obstacles to ELL children's successful school achievement may include low levels of family income and education, lack of familiarity with recommended parenting practices in the United States, poor English proficiency, and low levels of attendance in early education programs (Palacios et al., 2008). Children of immigrants who do not attend preschool are less likely to pass tests of English oral language proficiency than peers who experience early childhood educational settings. First-generation Black and Latino children encounter many obstacles associated with minority status in the United States, including discrimination, racism, and spatial segregation, all of which are risk factors for educational achievement. Over time and through subsequent generations, experiences with discrimination, low-achieving schools, and poor employment opportunities, compounded by the potential loss of protective traditional cultural factors, may translate to lower levels of overall achievement.

ELL student achievement is also greatly influenced by whether a student is first-generation (born in a foreign country), second-generation (born to foreign-born parents), or third-generation (Leventhal et al., 2006). One study (Palacios et al., 2008) examined the reading achievement trajectories of immigrant children from kindergarten to third grade, as indexed by generational status. First- and second-generation children had higher achievement scores by spring of kindergarten than did third-generation children. In addition, first-generation children grew in reading achievement at a faster rate than did third-generation children.

Because Mexican-Americans are one of the fastest growing immigrant groups in the country, there is a multitude of research documenting their academic trajectory in particular. Mexican-American immigrant children are more likely to be poor and to attend lower quality and more segregated schools than nonimmigrant children (Leventhal et al., 2006). Rates of enrollment in high school were almost five times lower for Mexican-American immigrants than nonimmigrants, and family socioeconomic characteristics only partially accounted for this discrepancy (Leventhal et al., 2006).

Legal Information

Educating students who are learning English begins with Lau vs. Nichols (1974), in which the U.S. Supreme Court recognized that ELL students would be locked out of the educational system unless schools developed instructional programs that would give them access to a meaningful education despite the language barrier (Guglielmi, 2008). Although the Court prescribed action, it gave local school districts the power to decide which particular educational programs to institute. As a result, the quality and quantity of

services and instruction vary greatly across states, districts, and schools (Callahan et al., 2008).

Despite evidence to the contrary, the importance of bilingual education is frequently questioned in today's school system, as has been demonstrated by new mandates for English-only instruction laws in states such as Arizona (Title 15, Chapter 7, Article 3, 2000) and California (Title 5, Chapter 11, Subchapter 4, 1998). New legislation under No Child Left Behind (NCLB) requires states to develop or adopt sound assessments to validly measure ELL students' English language proficiency (ELP). The legislation requires states to develop ELP standards, to annually measure ELL students' attainment of ELP based on those standards, and to set annual measurable achievement objectives (AMAOs) for ELL students as part of school accountability systems (Wolf, Farnsworth, & Herman, 2008).

Theories of Language Acquisition

Much research (Cardenas-Hagen et al., 2007) has been done about cross-linguistic transfer and how it might pertain to bilingual education programs and immigrant student success. Cross-linguistic transfer occurs when students learning another language have access to and use linguistic resources from their L1. Research on monolingual speakers in many languages affirms that phonological awareness is a precursor for learning to read. The cross-linguistic transfer of foundational skills such as phonological awareness and the alphabetic principle may illuminate the role of language on literacy acquisition for ELLs.

The theory behind bilingual education programs is that students learning to read in L2 benefit from their L1 reading knowledge because skills in L1 have components that transfer to learning to reading L2. One study (Cardenas-Hagen et al., 2007) found that although phonological awareness in English was most strongly related to English language proficiency, Spanish language proficiency and Spanish phonological awareness skills also contributed significantly to the variance in English phonological awareness. Phonological awareness refers to the ability to identify and manipulate units of speech (e.g., syllables, phonemes, onsets, rimes), and the connection between phonology and letter sound correspondence provides the initial foundation on which reading, writing, and spelling develops.

The study examined the effects of initial first and second language proficiencies, as well as the language of instruction that a student receives, on the relationship between native language ability of ELL students and their development of early literacy skills and the second language. The study found that Spanish-speaking students with high Spanish letter name and sound knowledge tend to show high levels of English letter name and sound knowledge; ELLs with low Spanish and English letter name and sound knowledge tend to show high levels of English letter name and sound knowledge when they are instructed in English. Letter name and sound identification skills are fairly highly positively correlated across languages in the beginning of the kindergarten year.

Phonological awareness skills appear to be the area with the most significant and direct transfer of knowledge, and language skills do not appear to be a factor in the development of phonological awareness. The relationship between oral language skills

across languages was low, suggesting little relationship between oral language skills across languages at the beginning of the kindergarten year. The language of instruction for Spanish-speaking ELLs may produce varying results for different students. Another study (Palacios et al., 2008) showed that the timing of second language acquisition and English proficiency may also be of critical importance to early developmental trajectories.

Types of Educational Programs for ELLs

Bilingual education theory allows that proficiency in one's native language (L1), however achieved, contributes to the development of second language (L2) literacy and ultimately to satisfactory academic achievement in ELL students, regardless of the particular instructional program in which they may have been placed (Guglielmi, 2008). Additive bilingualism (i.e. the development of fluency in both L1 and L2) has been associated with a variety of general cognitive advantages in the areas of divergent thinking, nonverbal reasoning, concept formation, metalinguistic awareness, creativity, and cognitive flexibility (Guglielmi, 2008).

In dual language programs, the goal is for students to maintain L1 while learning L2. The language of instruction is both L1 and L2, with an emphasis on maintaining L1 while acquiring L2. This program reflects the theory that it takes five to seven years for cognitive, academic learning, and that students can transfer skills and knowledge from one language to another (Cardenas-Hagen et al., 2007).

As many as 50% of ELLs in the United States receive reading instruction in their primary language (Cardenas-Hagen et al., 2007). This instruction is premised on

theoretical frameworks that assert that Spanish-speaking ELLs who learn to read in L1 may position themselves for higher levels of literacy in both L1 and L2. Schools with no language majority were more likely to provide instruction in English (Cardenas-Hagen et al., 2007).

One longitudinal study (Callahan et al., 2008) evaluated the efficacy of ESL programs with Mexican-American students. The researchers looked at the effect of ESL placement on Mexican-American students' achievement for first-, second-, and third-generation adolescents separately in schools with many and few immigrant students. The results were that the effect of ESL placement varies by both immigrant concentration in the school and by students' generational status. ESL enrollment may be protective for second-generation Mexican-American adolescents in high immigrant concentration schools, and may prove detrimental for first-generation adolescents in contexts with few other immigrant students.

Schools with many immigrant students may offer services and a climate better suited to the particular academic and social needs of immigrants than schools with relatively few immigrant students. Alternately, if a higher concentration of immigrants in a community leads to lower-quality schools, fewer academic opportunities, or the possibility of social marginalization, then immigrant students may be better served in schools serving predominantly nonimmigrant populations.

Fifteen states experienced more than 200% increase in students requiring ESL services between 1992 and 2002. In highly competitive, majority-White environments, Mexican-American students fare poorly, and they experience high levels of social and

academic marginalization. Mexican-American students have been found to complete less math and science coursework than their White or Asian peers.

Another study (Cardenas-Hagen et al., 2007) examined the effect of bilingual education with immigrant students. The study looked at both English language immersion, where language arts are conducted primarily in English and transitional bilingual programs, and where L1 is used in kindergarten and first grade, with progressive shifts to English. In addition, the study looked at dual language programs, where students are instructed equally in English and L1 in language arts courses. The study found that early Spanish skills predicted English outcomes at the end of kindergarten, after controlling for early English skills. It appears that knowledge of Spanish letter name and sound identification skills is transferred. Early Spanish skills predicted later English skills after controlling for early English skills. There is evidence that providing explicit instruction in foundational skills (phonological awareness, phonics) in L1 may assist ELLs in transitioning to reading in L2.

Measuring English Language Proficiency

Recently, states have had to make rapid policy changes in order to comply with the NCLB requirement to measure the progress of ELL students' English language attainment (Wolf et al., 2008). The NCLB legislation stipulates that the constructs of an ELP assessment should correspond with the state's ELP and content standards to measure the progress of appropriate English language development. However, there are currently many differences between state ELP standards and the assessments, including terminology and descriptions of each level of ELP.

One study (Wolf et al., 2008) looked at ELP assessments used in the 2006-2007 school year. At that time, 43 states had been using their ELP assessment for less than five years; eleven of those states introduced a new ELP assessment during the 2006-2007 school year. Almost all states utilized their ELP assessments for multiple purposes, such as identifying ELL students, determining their level of proficiency, and placing students into appropriate academic programs.

Indeed, one of the common purposes of an ELP assessment is to determine when students have sufficient proficiency to be successful in English-only classrooms and thus are ready to exit ELL status. Previous research on the academic achievement of students who were transferred into English-only classrooms yielded mixed results in terms of the students' subsequent academic performance relative to non-ELL students. If a state assessment does not accurately reveal individual students' English proficiency level, the individuals may be placed in inappropriate academic environments, which may impede their subsequent academic progress.

One of the most common criticisms of traditional ELP assessments has been that their constructs primarily focused on social language rather than students' ability to perform in an academic setting. Newly developed ELP assessments are intended to measure both academic and social language. However, even though all states mentioned "academic English" in their standards, there are a plethora of definitions of academic English proficiency in the states' ELP standards and most of these definitions were vague. For example, academic English constructs were described by listing tasks that occur in academic settings without clarifying what aspects of language ability were

involved these tasks. One state described that the reading section of the state's ELP assessment assessed whether a student "interprets hypotheses," "supports opinion and conjectures," is able to "read between the lines," and so on, without clarifying with what linguistic resources the student is or is not able to do this.

Another incongruence with ELP assessments is that there is wide variation in the ways in which states report scores and calculate them. Among the ELP assessments reviewed in the study, there were only 13 publicly available validity studies. This is partially due to changing conceptions towards the meaning of academic English and the lack of a commonly accepted form of measurement.

The studies suggest that more research is needed regarding how students acquire academic language; for example, it is well known that second-language learners tend to master specific grammatical features in a well-defined sequence, especially in the early stages.

English Language Learners with Learning Disabilities

Description of ELLs with LD

The recent reauthorization of the Individuals with Disabilities in Education or IDEA (2004), which is a law ensuring services to children with disabilities, requires states to develop policies and procedures to prevent over-identification of children by race or ethnicity in special education programs. The difficulty with carrying out this law is that ELL students share some characteristics with LD students, namely average cognitive skills, but low achievement in the classroom. Ortiz et al. (2006) defines an ELL student as a student who has such limited English skills that they cannot benefit

from general education instruction provided entirely in English without a special language program support. This is a similar definition to an LD student, who cannot benefit from general education without special education support. Because ELLs share this characteristic of having trouble in general education with LD students, they are particularly vulnerable to falsely positive identification of an LD (Ortiz, Wilkinson, Robertson-Courtney, & Kushner, 2006). The resulting outcome of a disproportionate representation of ELL students in special education has been a major concern in the field (Haager, 2007).

Assessment results across the United States show that ELL students do not achieve academically to the same degree as non-ELL students. Several studies have found significant gaps between the performance of ELL students and other students on reading and mathematics measures (Haager, 2007; Green and Ortiz et al., 2006). One study (Haager, 2007) found that over 50% of ELL students score in the bottom third in reading or mathematics. The number of ELL students that passed all parts of the 2001 Texas statewide achievement test ranged from 66% for third graders to 33% for tenth graders (Texas Education Agency, 2001). Overall, the 2004 statewide passing on assessments for ELL students was 51% for reading and language arts, 48% for math, and 72% for writing (Texas Education Agency, 2004). The results of a 1996 National Assessment of Educational Progress (NAEP) revealed that 41% of ELLs were scoring one year below and 29% were scoring two or more years below grade level on tests of reading and language arts. Clearly, this discrepancy in achievement is the symptom of an underlying problem in how ELLs are both tested and educated.

ELLs that are native Spanish speakers have the lowest achievement rates of all the groups of ELLs. This is particularly troubling because while there are special language programs provided for more than 400 language groups, nearly 80% of ELLs are Spanish speakers (Ortiz et al., 2006). One problem is that students who speak Spanish at home are more likely to have repeated a grade than students from other language groups (Ortiz et al., 2006). Another issue, is that according to the 2005 NAEP, only 13% of fourth-grade Hispanic students and 15% of eighth-grade Hispanic students meet proficiency reading standards (U.S. Department of Education, 2005). Additionally, a study by Wilkinson and Ortiz (1986) found that after three years of special education placement, Hispanic students had lower verbal and performance IQ scores than they had before entry into special education. This indicates that they are not receiving appropriate aid in special education. In fact, all of these numbers indicate that Hispanic students in particular are not receiving a free and appropriate education as guaranteed by the United States government.

Discrepancy Testing with ELLs

One reason for over-representation of ELL students in special education is that the tools used to measure LDs are not appropriate for this population (Hardin, Roach-Scott, & Peisner-Feinberg, 2007). Frequently, children who are culturally and linguistically diverse fail initial developmental screenings (Hardin et al., 2007) and as a result, many are inappropriately placed in special education. Some studies (Lopez et al., 1997) have found that ELLs that are labeled as LD were almost exclusively tested in English. Professionals responsible for identifying students for special education may not be aware

that ELLs may be able to orally communicate in English in social situations in as little as one or two years, but the skills needed to be cognitively and academically proficient may take five to eight years to develop (Hardin et al., 2007). There are also few, if any, satisfactory tools available to measure English proficiency, which makes it difficult to determine what degree of English Language proficiency is necessary for a cognitive and achievement test to be valid (Hardin et al., 2007).

For ELL students, the language in which tests are administered is the main source of test invalidity (Solano-Flores & Li, M, 2008). The linguistic demands of a test simply do not replicate across tests when translated. When given to ELL children, the scores may reflect English language knowledge instead of cognitive ability (Lopez et al., 1997). Even within the same language, different dialect versions of the same item, in combination with the translators' skills, may produce different sets of linguistic demands for the same item. One study (Solano-Flores & Li, M, 2008) gave ELLs the same set of items in English and their native language and examined the measurement error due to language and its interaction with student, item, and rater. The study also looked at dialect to see whether subtle linguistic differences in wording and word use could produce differences in the performances of ELL students depending on whether they were tested with a standard version of their native language or versions that were sensitive to the local dialects of their native language. Dialect accounted for 20% of the variance, indicating that dialect can be as important as language as a source of measurement error when testing ELLs.

RTI for ELL Students

Does RTI provide a similar ability to identify students with an LD as it does with native English speaking students? Some studies ((Mathes et al., 2007; Kamps et al., 2007; Ortiz et al., 2006) have begun to examine the generalizability of utilizing RTI in determining eligibility of ELL students for special education. It appears that RTI lessens the likelihood of special education placement for ELL students, though it is unclear if it is to the same extent as English as a first language (EFL) students. However, several studies do show that RTI does increase achievement in ELL students. One study (Mathes et al., 2007) found that students receiving tier 2 instruction, regardless of whether it was provided in Spanish or English, did better than students who were provided with core instruction alone. The results of another study (Kamps et al., 2007) indicated greater outcomes for ELL students participating in tier 2 interventions using curricula with a direct instruction approach and delivered in small groups than those enrolled in a comparison group. The study indicated no differences for ELL and English-only students receiving second tier interventions. Between 50-60% of the ELL students in direct-instruction interventions group were at benchmark at the end of first grade as compared to the comparison group, only 17% of which met benchmark criteria. This indicates that the RTI does increase the academic achievement of ELL students, but more research is needed.

RTI uses curriculum-based measures (CBMs) for both screening and progress monitoring. It helps determine the students' progress, and whether they should proceed to a higher tier for more intensive instruction. Some evidence (Haager, 2007) suggests

that CBM processes are useful with ELL students. First grade ELL students made similar progress to non-ELLs using nonsense word reading and oral reading fluency measures. However, false positives are a concern in when using these tools with ELL students, because they have not been adequately included in the norming samples.

Proposed Research Study

Statement of Problem

Overrepresentation of ELL students in special education has been major concern in the field. Studies have found that students who are ELL were 27% more likely than children who speak English as their native language to be placed in special education in elementary grades, and almost twice as likely to be placed in special education in middle school and high school (Rueda & Windmueller, 2006). In particular, ELL students are overrepresented in their diagnosis of a learning disability in reading. Other data has found that 56% of students who are ELL in special education have a reading LD (Klinger et al, 2008). This indicates that the current methods of evaluating ELLs for a reading LD may not distinguish between a language acquisition issue and a learning disability in reading.

In general, there are two methods for establishing the presence of a learning disability: RTI and the Discrepancy Method of Identification. The discrepancy method uses normed intelligence tests to test for a discrepancy between cognitive abilities and achievement. However, there are several issues with using tests of cognitive abilities with an ELL population. One problem is that these tests may not generalize to ELL students, because they are not included in the norming sample. Also much is "lost in translation" when students who are not native English speakers are tested in English. This results in, as one study found, underestimating the potential of linguistically diverse students (Klinger et al, 2006). While accommodations, such as use of a dictionary or glossary, or use of simplified language, have been found to be moderately helpful (Abedi,

2006), most students are tested in English without accommodations (Klinger et al, 2006). Ochoa et al (1996) found that only 1% of school psychologists attempted to determine if a discrepancy between cognitive ability and achievement occurred both in English and a student's native language. Given these issues, it is clear that those making special education placement decisions based on these tests may not be treating the results with appropriate skepticism.

This leaves the other option of testing for a learning disability which is RTI. The major shortcoming with this method is that there may not be enough existing evidence-based interventions specifically for ELL students to properly conduct the RTI process. Additionally, some researchers are concerned with false positives (Haager, 2007) when using interventions that have not used ELLs as part of the normed sampling and development process. Another problem with RTI, in general, is that it may not be the best method to distinguish between low achievement and an achievement discrepancy. Some argue that RTI measures only low achievement. A major benefit, however, of RTI is that it would require interventions to be attempted before a special education referral. Teachers often refer ELL students for learning disability without attempting an intervention to determine if the learning problem is caused by not understanding language.

Purpose

The purpose of the proposed study is to see how the two methods of learning disability identification (RTI and discrepancy testing) differ in proportion of ELL students they identify with a reading LD. In other words, does it matter which method is

used? A difference would indicate that the two methods do not measure the same construct within the ELL population. Additionally, the second stage of the study will see how the proportions of LD diagnoses identified by the two methods vary based on English language proficiency. The results will show if language proficiency has an effect on the number of reading LD diagnoses made with each test. It will also show the relative strength of that difference. The purpose is to understand the importance of language proficiency on the results of the two different methods of testing. Knowing which method is less affected by language proficiency is important information for school psychologists who are making special education placement decisions.

Research Questions and Hypotheses

Research Question 1

Do RTI and discrepancy testing identify the same proportion of ELL students with a reading LD?

Hypothesis 1

It is hypothesized that the proportion of discrepancy testing identifying a reading LD significantly different than the proportion of RTI identifying a reading disability.

Rationale. The use of RTI, especially if the intervention is designed to support students' cultural and linguistic diversity, can make sure that students receive quality instruction before making an LD determination. Research has found that early interventions that combine phonological awareness and ESL strategies are promising (Klinger et al, 2006). Previous research of ELL students who were instructed with the Spanish curriculum Lectura Proactiva (Mathes et al, 2003), which is based in current research of the transfer

of reading skills between Spanish and English, were significantly more likely to show posttest differences in Letter-Sound Identification, phonological awareness, passage comprehension, and reading fluency. This is important, because several studies have found that Spanish word recognition and phonological awareness are better predictors of English pseudoword and word reading than English word recognition (Klinger et al, 2006; Vaughn et al, 2006). In other, proficient Spanish readers are able to transfer those skills into English. In general, several studies support the RTI model for ELLs (Healy et al, 2005; Lovett et al, 2008). Discrepancy testing may not do a good job of distinguishing between students with an LD and students who are ELL. Students with LDs and students who are ELLs share similar characteristics, such as difficulty with test items that have unfamiliar words or a complex linguistic structure (Abedi, 2006). Students may do poorly on the tests as a result of English proficiency as opposed to lack of content knowledge.

Research Question 2

Do RTI and discrepancy testing identify reading LD diagnoses at the same proportion for ELL students as for EFL students?

Hypothesis 2

It is hypothesized that RTI and discrepancy testing will identify a larger proportion of ELL students as possessing an reading LD than EFL students.

Rationale. ELL students would be more likely to be diagnosed with an LD than EFL students. Most researchers believe that particularly for students who have extremely limited English proficiency, the results from tests used for assessment may be misleading

(Abedi, 2006). Tests used for ELL students have lower reliability and validity (Abedi, 2006). Norming and standardization procedures tend to exclude children from ELL backgrounds (Lopez et al, 1997). Studies (Lopez et al, 1997) have shown that when cognitive measures are given in English they instead become a language literacy test. Many tests may be unnecessarily linguistically complex, which is a source of measurement error, and is considered a construct-irrelevant factor in assessment (Abedi, 2006). However, studies have shown that reducing the linguistic complexity of items may close the performance gap between ELL students and EFL students (Abedi, 2006). This may be a better option than when tests are translated. When tests are translated they lack validation in the new language, as they may not measure the same set of characteristics or skills (Lopez et al, 1997). Nonverbal tests have weak technical data and none have been normed with the ELL population (Lopez et al, 1997). One study (Kamps & Greenwood, 2005), which looked at several selected RTI curricula among first grade teachers who employed the three-tiered RTI model, found that while there was growth in the experimental group that received RTI, students with disabilities, behavioral risks, and students who were ELLs did not show growth. However, another study (Lovett et al, 2008) found that both EFL and ELL students did better than the control groups, and there were not statistically significant differences in reading growth in the posttests of the intervention.

Method

Participants

Participants will be recruited from Austin Independent School District (AISD), which is a large school district in central Texas, in which 23% of the students are ELL and 96% receive Bilingual Education (BE) or English as a Second Language (ESL) services. Participants will be 264 ELL students in third grade bilingual education classes and 44 EFL students in monolingual education classes. Exclusionary criteria will include a previous diagnose of mental retardation and participation in special education.

AISD defines English-Language-Learners as students who (1) were not born in the United States or whose native language is a language other than English; or (2) come from environments where a language other than English is dominant; or (3) are American Indians and Alaskan Natives and who come from environments where a language other than English has had a significant impact on their level of English proficiency; and who, by reason thereof, have sufficient difficulty speaking, reading, writing, or understanding the English language, to deny such individuals the opportunity to learn successfully in classrooms where the language of instruction is English or to participate fully in our society.

Instrumentation

**Monitoring Basic Skills Progress – Second Edition* (Fuchs, L., Hamlett C., & Fuchs, D., 1999). This monitors students acquisition of basic skills in academic areas, such as reading, using curriculum-based measurements. The program is computer administered and scored. The students take short tests (on a weekly or bi-weekly basis) that are

selected based on previous performance. The student sits in front of the computer and is exposed to a reading passage with blanks. The student uses the 'mouse' to click a 'blank' and three different words are provided as options with each click on the blank. When the student is satisfied with the response, he or she moves to the next blank. The material is exposed for 2 1/2 minutes and then the computer screen changes and the scores for the student are presented. There are 30 passages at seven grade levels. The difficulty of the reading passages is based on Fry's readability formula. The results are graphed, enabling the student and teacher to visually monitor their progress. Additionally, the manual assists the teacher when alternative-teaching approaches may be needed to be utilized. Measures of reliability were reported to be satisfactory, ranging from .73 to .99. Student ability to record responses independently was also found to be accurate. Expert opinion of content was used to assess validity and high agreement was noted.

**The Woodcock Johnson Tests of Cognitive Abilities and Achievement* (Woodcock, R., McGrew, K., Mather, N., Schrank, F., 2001). The WJ III Tests of Cognitive Abilities are explicitly designed to assess a student's abilities on many Cattell-Horn-Carroll/Gf-Gc (CHC) cognitive abilities, not just a total score or a few abilities. Each of seven tests in the Standard Battery is designed to measure one broad ability. The Extended Battery offers seven more tests to so that two narrow abilities are measured for each CHC broad ability. There are three Standard and three Extended tests that contribute to additional Clinical Clusters. Tests can also be combined into a General Intellectual Ability (GIA Std) score of 7 or 14 tests (GIA Ext) and into several Cognitive Categories. Examiners are permitted to select the tests they need to assess abilities in which they are interested

for a particular student. The WJ III Cog was normed on 8,818 children and adults (4,783 in grades kindergarten through 12) in a well-designed, national sample. The same persons also provided norms for the WJ III tests of academic achievement (WJ III Ach), so the ability and achievement tests can be compared directly, and cognitive and achievement tests can be combined to measure CHC abilities. Research provides strongly supports that the Woodcock Johnson-III measures CHC abilities well (Floyd, Evans, McGrew, 2003; McGrew 2005; Schrank, Flanagan Woodcock, & Mascolo, 2002). The current study will use subsets from the WJ-III Cog and the WJ-III Ach to assess the reading abilities of students. These include subtests 1-7 and 11-17 on the WJ III Cog, and Letter-Word Identification, Word Attack, Passage Comprehension, Reading Fluency, and Reading Vocabulary from the WJ III Ach.

* *The Woodcock-Munoz Tests of Cognitive Ability- Revised in Spanish* (Woodcock & Munez-Sandoval, 1996). The Bateria is based on the Cattell-Horn-Carroll (CHC) model of cognitive abilities, a theory which focuses on measures of multiple broad abilities (i.e. fluid intelligence, crystallized intelligence, short-term acquisition and retrieval, cognitive processing speed, correct decision speed, and quantitative knowledge.) The Bateria does not use norms based on Spanish-speaking populations in the United States. Instead, the test developers used a sophisticated statistical procedure that allowed the test to develop a pool of items in Spanish that were of similar difficulty to the items in the English version of the Woodcock-Johnson Tests of Cognitive Ability-Revised (Woodcock & Johnson, 1989.) The difficulty levels of the items in the Spanish Bateria where calculated by

testing approximately 2,0000 native Spanish-speaking individuals from Costa Rica, Mexico, Peru, Puerto Rico, Spain and the United States. These results were used to 'calibrate' the scale, that is, to equate it statistically to the English-language WJ III. First, items on both the WJ III and the Bateria III were scaled according to their item difficulty. Calibration statistics were then examined for equivalent items, and Bateria III item difficulties were statistically adjusted to the scale of the WJ III. Using these comparisons, each test in the Bateria III was scaled or equated to their corresponding difficulty in the WJ III. The test developers claim that the procedure is innovative and useful because it allows examiners to compare equitable scores across two tests administer in two different languages. However, comparing Bateria III performance to WJ-III norms is controversial. It is thought to be the best available Spanish-language measure of cognition and achievement. The corresponding subtests will be used as for the Woodcock Johnson-III tests of achievement and cognition (in English.) This includes Habilidad intellectual general-Escala Extendida, which are subtests 1-7 and 11-17 to measure general cognitive ability. Also, to measure achievement, the following tests will be given: Identificacion de letras y palabras, analisis de palabras, comprension de textos, fluidez en la lectura, and vocabulario de lectura.

Procedure

Approval by Human Subjects Committee

The study will be conducted in compliance with the ethical standards designated by the American Psychological Association, as well as the standards set forth by the University

of Texas at Austin. Before the study, researcher will obtain approval from the Departmental Review Committee with the Department of Educational Psychology and the Institutional Review Board of the University of Texas at Austin.

Approval by the school district

The researcher will provide a written proposal to the external research coordinator of Austin Independent School District requesting approval to conduct research on selected campuses. Once permission is granted, the researcher will obtain permission from principals of selected campuses to conduct research. Then the researcher will meet with school faculty and staff to discuss the designated responsibilities and requirements of the proposed research study.

Recruitment of participants

After permission is granted from school principals, third grade teachers of both bilingual and English-only classrooms will be asked to implement a RTI reading curriculum in their classrooms. Each teacher will report to the researcher the number of students who participate in each tier of RTI. Because of the time commitment required to complete the RTI, all participating teachers will be entered into a drawing for a $100 gift certificate to be awarded to one teacher in recognition for his or her efforts to assist in research.

Bilingual teachers at participating campuses will teach an RTI curriculum for Spanish speakers called *Enhanced Proactive Reading* (Vaughn, S., Cirino, P., Linan-Thompson, S., Mathes, P., Carlson, C, Cardenas-Hagan, E., et al (2006). This intervention met standards with reservations according to the What Works Clearing

House (US Department of Education, 2007). The review focused on outcome of reading achievement and English language development. The review found it to have potentially positive effects, or "evidence of a positive effect with no overriding contrary evidence." Two studies have been conducted on this program. A sample size of 8 schools and 131 students was used. The study included students in grades 2 to 5. The studies had statistically significant positive findings at the .05 alpha level. A study of the program by Vaughn, Cirino, et al., (2006) used a randomized controlled trial to not be statistically significant but substantively important for reading achievement. Another study by Vaughn, Matthes, et al., (2006) did find the intervention to be statistically significant and substantively important for reading achievement. Neither study found English language development to be statistically significant or substantively important.

Teachers in English-only classrooms in select campuses will teach *Success for All* (Slavin, Madden in Conjunction with Johns Hopkins University). One study met the What Works Clearinghouse (WWC) evidence standards and six studies met What Works Clearing house evidence standards with reservations. The studies were conducted in more than 90 elementary schools with 6,000 students. The study had medium to large effect sizes for alphabetic, comprehension, and general reading achievement. None of the studies met WWC standards for fluency. One study (Borman et al, 2006) did a three-year longitudinal study of 1,425 students at 18 intervention schools and 17 comparison schools. They examined scores on the Woodcock Reading Mastery Test and found statistically significant positive effects for Word Identification and Word Attack subtests.

They also found statistically significant effect on comprehension. The intervention was also found to substantively important, in that it had an effect size greater than .25.

The teachers receive 12 hours of professional development in how to implement the different reading interventions. They will then receive 6 hours of additional professional development six weeks after instruction begins. The staff will also meet weekly for about two hours with the study developer to view videotaped lessons, discuss issues regarding implementation, and problem solve in how to accelerate growth for specific students.

Then, students in all of the participating classrooms will be invited to participate in additional discrepancy testing at the end of the school year, after having completed the RTI curriculum. Each student will receive a letter describing the nature of the proposed study, a student assent form (see Appendix C), and a parental consent form (see Appendix A and B). The assent and consent forms will describe the nature and rationale for the study and outline any potential risks or benefits of participation. Each form will include a section explaining that study participation is voluntary, that refusal to participate will not impact the student's or parents' relationship with the school, and that assent or consent may be revoked at any time. Spanish and English forms will be available. School personnel at each campus will collect completed assent/consent forms and give them to the researcher. All participants will receive a University of Texas at Austin bracelet in recognition of their participation.

Collection of Data

Students will be measured every three months to determine their progress in meeting reading benchmarks. Students that fall below the 25th percentile will be referred for Tier 2 instruction. Students that after three months continue to fall below the 25th percentile will be referred for Tier 3 instruction. This time length was determined by guidelines from the National Research Center on Learning Disabilities which says "although no clear consensus exists on the duration of Tier 2 and Beyond intervention, in general, the research supports 10 to 12 weeks for each round of information."

At the end of the intervention, graduate students trained in assessment will individually administer the cognitive battery of selected subtests from the WJ-III Cog and WJ-III Ach to each participant. The testers will remove each participant individually form the classroom and conduct the assessment in a private location in the school. Students will be pulled from classrooms at times that have been previously approved by teachers and administrators.

Once the cognitive battery has been given, the participants' scores on the WJ-III will be calculated using the WJ-III Compuscore and Profiles program (Schrank & Woodcock, 2001.)

Data Analyses and Expected Results

Preliminary Analysis:

A power analysis was conducted using G-POWER software, version 3.0, to determine an appropriate number of participants needed for the present study to obtain statistical significance with a medium effect size with 80% power. It was determined that a sample size of 264 achieves 80% power with a significance level (alpha) of .05 to perform McNemar's chi-square test. For the second part of the study, a power analysis was conducted for chi-square goodness-of-fit contingency table to determine the sample size that would achieve an 80% power with an alpha of .05 and a medium effect size. The total sample size needed was 88. In order to ensure equal n values for both the ELL and EFL groups, 44 students for each groups will be used in the model.

Test of Hypothesis

Hypothesis 1:

It is hypothesized that the proportion of discrepancy testing identifying a reading LD significantly different than the proportion of RTI identifying a reading disability.

Data for the first part of the from the study will be analyzed using McNemar's chi-square test. This test was chosen because it tests the homogeneity of proportions in a sample that is measured twice. The McNemar test is applied to a 2x2 table. The following table summarizes the agreement between the methods of reading LD testing with presence of an LD. It does this by testing for marginal homogeneity, or how well the rows equal the corresponding columns.

		Diagnosis with RTI	
Diagnosis with Discrepancy Testing		YES	NO
	YES		
	NO		

It is expected that difference betweens the rows and the columns will be statistically different. Therefore the proportion of LD outcomes using the two different methods will be different.

Hypothesis 2

It is hypothesized that RTI and discrepancy testing will identify a larger proportion of ELL students as possessing an reading LD than EFL students.

A multiway contingency table will be constructed and the association between EFL students and ELL students will be determined by using a conditional odds ratio. The odds ratio is a way of comparing whether the probability of a certain event is the same for two groups. An odds ratio of 1 implies that the event is equally likely in both groups. An odds ratio greater than one implies that the event is more likely in the EFL group. An odds ratio less than one means that the event is less likely in the EFL group.

Odds ratios also do not require variables be normally distributed or that relationships between variables be homoscedastic. That is, using odds ratios as measures of strength involves much more relaxed data assumptions as effect size measures. Data from both groups will be entered into a multiway contingency table as illustrated below:

		LD present	
English Language Proficiency	Reading LD identification	Yes	No

EFL	RTI	a	b
	Discrepancy	c	d
ELL	RTI	e	f
	Discrepancy	g	h

The pattern of association among the variables will be described using the conditional odds ratio. The association between Reading LD identification and LD present will be conditioned on the categories of English Language Proficiency. The part of the table in which the English Language Proficiency is EFL is a 2x2 table, from which the odds ratio can be computed. This is the conditional odds ratio of reading LD identification and LD present, conditioned on English Language Proficiency= EFL. This correlational odds ratio will be denoted COR (E, R/L=EFL). It's value is (a*d)/(c*b). Similarly, for the other four cells in the table COR (E, R/L=ELL) = (e*h)/(g*f). The ratio of the two conditional odds:

$$\frac{\text{COR (E, R/L=EFL)}}{\text{COR (E, R/L=ELL)}}$$

This equation expresses relative strength of the above effect for those who are EFLs as compared to those who are ELLs. It is expected that the students will be more likely to be found to have a reading LD when they are ELLs than when they are EFLs (i.e., the ratio will be greater than 1.)

Discussion

Summary and Limitations

Several limitations are present in this study. One limitation is that the study only looked at Spanish-speaking students. This is because there are not any reliable and valid assessment procedures available in other languages. However, the ELL population in the United States is an extremely heterogeneous population, and they effectiveness of the type of assessment used to measure an LD may differ based on the native language of the student. Additionally, there are several different dialects of Spanish. This means that assessment and instructional methods that use standardized Spanish may exploit different cognitive abilities than what is being tested.

Another limitation of the study is that there is no reliable and valid way of determining English Language proficiency. The tests that are available do not reflect current linguistic research in how English is learned (i.e. there tends to be a sequence in how different grammatical structures are mastered.) Therefore, while the study grouped all ELL students together, there may, in fact, be a wide spectrum of English proficiency that would impact the effectiveness of different types of ELL assessment procedures.

Yet another limitation is that because the students will be recruited from the same school district, schools, and classrooms, group effects may be present from sampling within a community. The major group effect would be teacher effects. While the intervention methods are standardized, each teacher has a different style of instruction that might affect the level of mastery from classroom to classroom. Other differences in

testing students by using RTI may include the noise-level in the classroom, seating arrangements, and additional directions.

There are also limitations of the statistical tests used. Categorical data, which was used in both parts of the study, do not provide the same depth of information as continuous variables. One limitation of the McNemar test is that it does not examine the extent of difference, only whether a difference has occurred. For the second part of the study, other statistical tests of association were considered. For example, one could use a loglinear model to analyze a multiway contingency table by either comparing more parsimonious models to a saturated model to test goodness-of-fit hypothesis. However, underlying assumptions, including independence, would have been violated, which is why using odds ratios was used to analyze the contingency table.

More research also needs to be conducted into evidence-based instruction for ELL students. There are few interventions that are designed for these students, and the amount of research available evaluating the studies that do exist is sparse. This makes it challenging to compare reading interventions for ELL students to interventions for native English speakers, for which there is an abundance of research.

Implications and Directions for Future Research

If the first hypothesis holds true, and one method of testing for and LD in the ELL population is more likely to find a reading LD diagnosis, then this indicates that it *does* matter which test is used to make special education determinations. Future research should confirm these findings. The implication of such a result (i.e. the rejection of the null hypothesis) is that RTI and discrepancy testing are measuring different constructs for

students who are ELLs. It is essential that school professionals who are making special education decisions to be aware that one test is more likely than another to positively identify at student for special education. This may cause school psychologists to be more aware of false positives for one method and wary of false negatives with another method.

If the second hypothesis is true, and there is an interaction between outcomes for ELL student and EFL student depending on the type of testing method used, then this implies that these tests are not very accurate with the ELL population. The magnitude of the interaction between these two populations will show the extent to which these tests are accurate. While certain factors may increase the likelihood of the presence of an LD (e.g. genetic factors) one's native language should not cause an increase in LD prevalence. A disparity between the numbers of ELLs and EFLs diagnosed with a reading LD indicates that this is more an issue with the type of test administered than the actual diagnosis of the student.

All of this information should help make school psychologists cognizant of the limitations and shortcomings of using either of these methods to test ELLs for placement into special education. There are extra considerations to be made when distinguishing between an ELL with an LD and one without, primarily because they *do* share so many characteristics with a student who has LD. Those making placement decisions must carefully consider if the problem is an LD or because the student merely has not had enough time to acquire proficiency in English. Both these scenarios contribute to lower achievement in general education. However, while the symptoms are similar, the cure is different. A student who is failing due to lack of English proficiency requires English

language instruction, while a student failing due to an LD requires special education intervention. In sum, it is essential that students who are ELL are not misplaced in special education, and are receiving the language instruction they need to succeed in the United States. Students who are ELLs have a right, under the law, to a free and appropriate education.

This proposed study is only the first step in determining how to distinguish between an LD and a language acquisition issue. The next step for future research is to measure what level of English language proficiency is necessary for the different methods of special education identification that have been discussed here to be accurate. It may be the case that with certain levels of English Language proficiency, these tests are almost as accurate as for native English speakers. However, it will not be until the creation of reliable and valid English Language Proficiency tests that we can understand how language confounds special education testing.

As previously mentioned, this study was limited to evaluating a reading LD in native Spanish speaking students. More research should be conducted in subjects other than reading. Most research concentrates on reading because it is has such a foundational role in education. However, students who are ELL have lower achievement than EFL students in math, science, and social science (Abedi, 2006). More interventions should be designed to instruct ELL students in these subjects, especially since taking algebra in eighth grade is strongly correlated with attending college (Callahan et al, 2008). Thus, the correct diagnosis of other specific LDs could have implications in terms of the successful academic trajectory of immigrant students.

Of course, more research attention is required towards students who have a native language other than Spanish. While Spanish speakers make up about 80% of the ELL population (Guglielmi, 2008), students who speak native languages that do not share cognates or other linguistic features with English may require different types of interventions. More research should be conducted to see which types of interventions are appropriate with different types of language groups. This would have especially large implications for those who are testing with RTI. The research would also serve to improve the quality of ESL services. Good instruction can help teachers the prevent making referrals that result in misdiagnosis.

References

Abedi, J. (2006). Psychometric issues in the ELL assessment and special education elgibility. *Teachers College Record, 108(11),* 2282-2303.

Austin Independent School District. *Bilingual Education (Data File.)* Retrieved from http://www.austinisd.org/academics/curriculum/subjects/bilingual/index.phtml.

Borman, G. D., Slavin, R. E., Cheung, A., Chamberlain, A., Madden, N., & Chambers, B. (2006). *Final reading outcomes of the national randomized field trial of Success for All.* Retrieved from Success for All website: http://www.successforall.net/_images/pdfs/Third_Year_Results_06.doc

Brooks, V., Hughes, C., Ruhl, K., Benedek-Wood, E., & Dexter, D. (2008). A content analysis of learning disabilities research & practice: 1991-2007. *Learning Disabilities Research & Practice, 23(4),* 184-193.

Callahan, R., Wilkinson, L., Muller, C. (2008, March). School context and the effect of ESL placement on mexican-origin adolescents' achievement. *Social Science Quarterly, Volume 89 (1),* 177-197.

Cardenas-Hagen, E., Carlson, C., Pollard-Durodola, S. (2007, July). The cross-linguistic transfer of early literacy skills: the role of initial L1 and L2 skills and the language of instruction. *Language, Speech, and Hearing Services in the Schools*, 249-259.

Cizek, G. (2004). review of the Woodcock Johnson® III. *Mental Measurements*

Yearbook,15.

Doll, B., & LeClair, C. (2004). review of the Woodcock-Munoz reading LD identification in ELL students 51, *Mental Measurements Yearbook, 17*.

Flanagan, D., Ortiz, S., Alfonso, V., & Dynda, A. (2006). integration of response to intervention and norm-referenced tests in learning disability identification: learning from the tower of babel. *Psychology in the Schools, 43(7),* 807-825

Flanagan, D., Ortiz, S., Alfonso V. (2007*). Essentials of Cross Battery Assessment (2nd Ed.)* Hoboken, New Jersey: John Wiley & Sons.

Fletcher, J., Denton, C., & Francis, D. (2005, November/December). Validity of alternative approaches for the identification of learning disabilities: operationalizing unexpected underachievement. *Journal of Learning Disabilities, 38(6),* 545-552.

Greasley, P. (2007). *Quantitative Data Analysis Using SPSS : An Introduction for Health and Social Studies.* Maidenhead: McGraw-Hill International (UK) Ltd.

Guglielmi, R. (2008). Native language proficiency, English literacy, academic achievement, and occupational attainment in limited-English-proficient students: a latent growth modeling perspective. *Journal of Educational Psychology, 100(2),* 322-342.

Hallahan, D., Keller, C., Martinez, E., Byrd E. S., Gelman, J., & Fan, X. (2007, Winter). How variables are interstate prevalence rates of learning disabilities and other

special education categories? A longitudinal comparison. *Exceptional Children, 73(2),* 136-146.

Han, W. (2008). The academic trajectories of children of immigrants and their school environments. *Developmental Psychology, 44(6),* 1572-1590

Hardin, B., Roach-Scott, M., & Peisner-Feingberg, E. (2007). Special education reading LD identification in ELL students referral, evaluation, and placement practices for preschool English language learners. *Journal of Research in Childhood Education, 22(1).*

Hollenbeck, A., (2007). From IDEA to implementation: a discussion of foundational and future responsiveness-to-intervention research. *Learning Disabilities Research & Practice, 22(2),* 137-146.

Huck, S. (2008). *Reading Statistics and Research (5th Ed.)* Boston: Pearson Education, Inc.

Kamps, D., Abbot, M., Greenwood, C., Arreaga-Mayer, C., Wills, H., Longstaff, J., Culpepper, M., Walton, C. (2007, Summer). Use of evidence-based, small-group reading instruction for English Language Learners in elementary grades: secondary-tier intervention. *Learning Disability Quarterly, 30,* 153-168.

Kavale, K., Kauffman, J., Bachmeier, R., & LeFever, G. (2008, Summer). Response-to-intervention: separating the rhetoric of self-congratulation from the reality of specific learning disability identification. *Learning Disability Quarterly,* 135-

150.

Keith, T. (2006). *Multiple Regression and Beyond.* Boston: Pearson Education, Inc.

Kim, S., Chao, R. (2009). Heritage language fluency, ethnic identity, and school effort of immigrant Chinese and Mexican adolescents. *Cultural Diversity and Ethnic Minority Psychology, 15(1),* 27-37.

Klinger, J., Alfredo, A., & Mendez Bartela, L. (2006, March/April). English Language Learners who struggle with reading: languague acquistion or LD? *Journal of Learning Disablities, 39,* 108-128.

Haager, D. (Summer, 2007). Promises and cautions regarding using response to reading LD identification in ELL students intervention with English Language Learners. *Learning Disability Quarterly, 30,*213-218.

Healy, K., Vanderwood, M., & Edelston, D. (2005). Early literacy interventions for English Language Learners: support for an RTI model. *The California School Psychologist, 10,* 53-63.

Leventhal, T., Xue, Y., Brooks-Gunn, J. (2006, September/October). Immigrant differences in school-age children's verbal trajectories: a look at four racial/ethnic groups. *Child Development, 77(5),* 1359-1374.

Liu, Y., Ortiz, A., Wilkinson, C., Roberson, P., & Kushner, M. (2008). From early childhood special education to special education resource rooms: identification, assessment, and eligibility determinations for English Language

Learners with reading-related disabilities. *Assessment for Effective Intervention, 33,* 177-187

Lopez, E., Lamar, D., & Scully-Demartini, D. (1997). The cognitive assessment of limited-English-proficient children: current problems and practical recommendations. *Cultural Diversity and Mental Health, 3(2),* 117-130.

Lovett, M., De Palma, M., Frijters, J., Steinbach, K., Temple, M., Benson, N., & Lacerenza, L. (2008). Interventions for reading difficulties: a comparison of response to intervention by ELL and EFL struggling readers. *Journal of Learning Disabilities, 41, 333*-352.

Mathes, P., Pollard-Durodola, S., Cardenas-Hagan, E., Linan-Vaughn, S., & Vaughn, S. (2007, July). Teaching struggling readers who are native Spanish speakers: what do we know? *Lanaguage, Speech, and Hearing Services in Schools, 38,* 260-271.

McCardle, P., Mele-McCarthy, J., Cutting, L., Leos, K., & D'Emilio, T. (2005). Learning disabilities in English Language Learners: identifying the issues. *Learning Disabilities in Research & Practice, 20(1),* 1-5.

McLellan, M. (2004). Review of monitoring basic skills progress—Second Edition. *Mental Measurements Yearbook, 14*.

National Research Center for Learning Disabilities. (2007, Winter.) *Research Based View of a Specific Learning Disability: Responsiveness to Intervention Overview*.

Retrieved from http://www.nrcld.org/topics/rti.html.

Ortiz, A., Wilkinson, C., Robertson-Courtney, P., & Kushner, M. (2006, January/February). Considerations in implementing intervention assistance teams to support English Language Learners. *Remedial and Special Education, 27(1),* 53-63.

Palacios, N., Guttmannova, K., Chase-Landsdale, L. (2008). Early reading achievement of children in immigrant families: is there an immigrant paradox? *Developmental Psychology, 44(5),* 1381-1351.

Rudas, T. (1998). *Odds Ratios in the Analysis of Contingency Tables.* Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-119. Thousand Oaks, CA: Sage.

Rueda, R. & Windmueller, M. (2006, March/April). English Language Learnes, LD, and reading LD identification in ELL students overrepresentation: a multiple-level analysis. *Journal of Learning Disabilities,39,* 99-107.

Russo, R. (2003). *Statistics for the Behavioural Sciences: An Introduction*. London : Psychology Press.

Sandoval, J. (2004). Review of the Woodcock-Johnson® III. *Mental Measurements Yearbook, 15.*

Shepard, L. (1980). An evaluation of the regression discrepancy method for identifying children with learning disabilities. *Journal of Special Education,*

14(1), 79-91.

Smithson, M. J. (1999). *Statistics with Confidence : An Introduction for Psychologists.* London : Sage Publications Ltd.

Solano-Flores, G., & Li, M. (2008, June). Examining the dependability of academic achievement measurement measures for English Language Learners. *Assessment for Effective Intervention, 33(3), 135*-144.

Spinelli, C. (2008). Addressing the issue of cultural and linguistic diversity and assessment: informal evaluation measures for English Language Learners. *Reading & Writing Quarterly, 24*, 101-118.

Turney, K., Kao, G. (2009, March/April). Barriers to school involvement: are immigrant parents disadvantaged? *The Journal of Educational Research, 102(4),* 257-271.

Van Emden, H. (2009). *Statistics for Terrified Biologists.* Chichester : John Wiley & Sons, Ltd.

Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Cardenas Hagan, E., Pollard-Durodola, S. D., Fletcher, J. M., & Francis, D. J. (2006). Effectiveness of a Spanish intervention and an English intervention for English Language Learners at risk for reading problems. *American Educational Research Journal, 43 (3),* 449–487

Vaughn, S. & Fuchs, L. (2003). Redefining learning disabilities as inadequate

response to instruction: the promise and potential problems. *Research & Practice – Special Series, 18(3*), 137-146

Vaughn, S., Linan-Thompson, S., Matthes, P., Cirino, P., Carlson, C., Pollard-Durodola, S., Cardenas-Hagan, E., & Frances, D. (2006, January/February). Effectiveness of Spanish intervention for first-grade English Language Learners at risk for reading difficulties. *Journal of Learning Disabilities, 39,* 56-73.

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., Cardenas-Hagan, E., & Franics, D. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *Elementary School Journal, 107 (2),* 153–180.

Wilkinson, C., Ortiz, A., Roberston, P., & Kushner, M. (2006, March/April). English Language Learners with reading-related LD: linking data from multiple sources to make eligibility determinations. *Journal of Learning Disabilities, 39(2),* 129-141.

Wolf, M., Farnsworth, T., Herman, J. (2008). Validity issues in assessing English Language Learners' language proficiency. *Educational Assessment, 13,* 30-107.