1-1-2009

Examination of Universal Screening in Response to Intervention at the Secondary Level

Debra Allison Layland
University of Denver

EXAMINATION OF UNIVERSAL SCREENING IN RESPONSE TO INTERVENTION AT THE SECONDARY LEVEL

_____

A Dissertation

presented to

the Morgridge College of Education

University of Denver

_____

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

_____

by

Debra Allison Layland

June, 2009

Advisor: Kent Seidel, Ph.D.

Author: Debra Allison Layland
Title: EXAMINATION OF UNIVERSAL SCREENING IN RESPONSE TO INTERVENTION AT THE SECONDARY LEVEL
Advisor: Kent Seidel, Ph.D.
Degree Date: June, 2009

ABSTRACT

The purpose of this study was to examine three means of universal screening to identify students in need of reading interventions at the secondary school level. Specifically the results of the Colorado State Assessment Program, the Test of Silent Contextual Reading Fluency (TOSCRF) (Hammill, Wiederholt, & Allen, 2006) and the Test of Silent Word Reading Fluency (TOSWRF) (Mather, Hammill, Allen, & Roberts, 2004) were used to group students in a Response to Intervention Framework of Tier 1 (needing no intervention), Tier 2 (needing intervention) and Tier 3 (needing intensive intervention). Categorization results were compared to determine the accuracy of identification by these screening tools. Analysis of the results indicated a significant difference between the CSAP and each of the other screening tools. Use of state assessment results alone resulted in under identification of students in need of interventions as compared to use of the TOSCRF or TOSWRF. No significant differences by gender across the three tests were seen. A significant difference was seen between the performance of Hispanic students and White students; this, however, could be attributed to the sample size. In addition, no significant interaction effects of gender by Hispanic/White ethnicity were noted for each of the three tests. Finally, no significance was seen within each screening tool by grade; there was, however significance across the three screening tools within each grade, with the exception of grade 9. Implications for practice, as well as future research, are discussed.

Table of Contents

List of Tables

List of Figures

Chapter One

Introduction

Sam is a twelfth grade student excited about his upcoming high school graduation. He works well with his hands, is a good student in math and plans to attend college the following fall. In contrast his performance in English continues to be poor. His mother has always been aware of his difficulties in reading, which originally surfaced in middle school. By the end of ninth grade Sam had failed the majority of his classes so he was placed in an alternative program for tenth grade. There he managed to complete his work; however, his difficulties in reading and writing persisted. Upon completion of tenth grade in the alternative program, Sam returned to general education and repeated English, only to fail again. His mother finally demanded a comprehensive evaluation, which indicated Sam had significant reading and spelling deficits. Recently, Sam took the ACCUPLACER, a placement test developed by the College Board (The College Board, 2005) to determine course placement of incoming students. The results indicate Sam would need to take remedial level English and writing before beginning college level classes. Sam does not have the skills necessary to successfully complete high school, let alone begin college level studies. Why is it that we have students like Sam who near completion of a public education program, yet are not proficient in basic reading and writing skills?

A reading crisis exists in the United States that will affect the nation's future. More than eight million students in fourth through twelfth grades read below grade level (Alliance for Excellent Education, 2006). The nation's measurement of academic

achievement, the National Assessment Education Program (NAEP, 2009), reading results are alarming. The average reading score for eighth graders has increased only one point since 2005 and only three points since 1992. The results are even more disturbing for low-income and minority students. Only 15% low-income eighth grade students read at a proficient level (Alliance for Excellent Education, 2006) and 89% of Hispanic and 86% of African American students in middle and high school read below grade level (National Center for Educational Statistics, 2004). In 2007, the gap between Caucasian and minority students in reading at the eighth grade level showed no change from 2005 results (Lee, Grigg, & Donahue, 2007). The crisis only worsens for students whose native language is not English. Four percent of English language learners (ELLs) and 20% of students classified as former ELLs were proficient or advanced on the NAEP in 2005 (Perie, Grigg & Donahue, 2005). In a typical high-poverty urban high school approximately half of the incoming freshman read two to three years below grade level (Balfanz, Herzog, & Mac Iver, 2007). Overall, experts in adolescent literacy estimate that as many as 70% of adolescents struggle with reading in some manner and are in need of learning strategies or interventions (Biancarosa & Snow, 2006).

In addition to the NAEP, each state has an assessment program to measure student progress in reading and math as part of accountability requirements in the No Child Left Behind Act of 2001 (Public Law 107-110). Each state is permitted to design and implement their basic skills assessment, and is required report results to the U.S. Department of Education. Colorado uses the Colorado Student Assessment Program (CSAP) to provide a picture of the progress its students are making toward academic standards, and to demonstrate the degree to which schools are ensuring students' learning

success (Colorado Department of Education [CDOE], 2009). Colorado is one of 30 states that recorded no significant change between 2005 and 2007 in reading scores at either fourth or eighth grade. CSAP results in 2005 and 2007 indicate 64% and 63% of eighth grade students performed at or above proficiency, respectively.

Despite the growing concerns over poor performance, instruction has been inconsistent for adolescents struggling in reading. Many secondary teachers do not believe they are responsible for teaching reading skills in their content area classes (Kamil et al., 2008). Middle School teachers often comment that if elementary teachers did a better job teaching reading, the problems at the secondary level would be solved or would not exist (Kamil, 2003). Students who display significant difficulties in reading are often referred for a special education comprehensive evaluation. Students who qualify as having a disability are then provided special education services to remediate their deficits. Unfortunately, special education has become an end in itself, rather than a means of providing direct, explicit, effective instruction to those that need it (President's Commission on Excellence in Special Education, 2002). Special education services at the secondary level can range from content area tutoring to alternative content area courses (Zigmond, 1990). Conderman and Petersen (2007) more recently noted that secondary special education services lack definition, are disjointed and uncoordinated, and invariably become tutoring programs that rarely include the type of instruction that leads to significant achievement gains.

One promising instructional framework is use in some districts and schools is Response to Intervention (RTI). The National Center on Response to Interventions (2008) states, "response to intervention integrates assessment and intervention within a multi-

level prevention system to maximize student achievement and to reduce behavior problems" (p. 1). With RTI, schools identify students at risk for poor learning outcomes, monitor student progress, provide evidence-based interventions and adjust the intensity and nature of those interventions depending on a student's responsiveness" (p. 1). RTI consists of four key components: universal screening, progress monitoring, levels or tiers of intervention, and fidelity of implementation. These components together provide a system of high quality research-based instruction, mechanisms for monitoring each student's performance as learning occurs, preventive interventions for those students at risk, alternative interventions to those students not responding to classroom or preventive interventions, and a measure of fidelity of implementation of interventions to ensure students had adequate opportunity to respond.

Universal screening is the first step in a RTI multi-tiered system. Such screening utilizes an efficient, low-cost, repeatable test of age-appropriate essential skills to identify students who are at risk for academic or behavior problems. Decision rules are based on "cut scores" that distinguish students who are at risk from students who are progressing adequately (Johnson, Mellard, & McKnight, 2006). Once identified, students at risk receive interventions that increase in intensity, frequency, and duration in order to improve their skills. However, if the screening process over identifies at-risk students valuable school resources may be misused, whereas a process that under identifies could result in some students not getting the critical assistance they need in order to succeed in school. Therefore, accurate universal screening is a critical step in a school's attempt to improve achievement for all students.

Statement of Problem

RTI has been successful in increasing student achievement in some elementary schools (Johnson et al., 2006), and may be a viable means to address the needs of secondary students who struggle in reading. Universal screening to identify students in need of intervention, the first step in an RTI framework, can be difficult in middle and high schools due, in part, to the lack of reliable and valid screening tools. To fill this void, some schools in Colorado use the results of the state assessment program, CSAP, to identify students at risk. However, the purpose of CSAP is "to provide a picture of how students in the state of Colorado are progressing toward meeting academic standards, and how schools are doing to ensure learning success of students" (CDOE, 2009, p. 1). Thus, screening to identify students in need of reading interventions may be an inappropriate use of the CSAP assessment instrument. In addition, reliance on state assessment results may, in fact, result in over or under identification of students at risk (Jenkins, 2003). As stated earlier, misidentification may either result in undue burden on a school's or district's resources or cause potential harm to a student that goes without needed assistance. Researchers and proponents of RTI recommend using specific screening assessments that are efficient, cost effective measures of specific skills (Batsche, Castillo, Dixon, & Forde, 2008; Chun & Witt, 2008; Fuchs & Fuchs, 2005; Johnson et al., 2006; Vaughn, 2005; Shinn, 2008; Sugai & Horner, 2007). When choosing a screening tool one must ask, "What is the most effective means of identifying students at risk for poor reading outcomes at the secondary level so interventions can be provided in an efficient and timely manner to ensure that all students graduate with the reading skills necessary for post secondary study or employment?" This question is extremely important given the

relatively limited time the secondary education system has with a student to address what could be a significant reading problem.

Purpose of Study

The purpose of this study is to examine three methods of universal screening to identify students in need of reading interventions at the secondary school level, and to compare the categorization results using the CSAP to the results of each of two published, nationally-normed screening assessments, the Test of Silent Contextual Reading Fluency (TOSCRF; Hammill, Wiederholt, & Allen, 2006) and the Test of Silent Word Reading Fluency (TOSWRF; Mather, Hammill, Allen, & Roberts, 2004). In addition, an analysis will determine if value is added to a CSAP-based screening process by adding the TOSCRF or TOSWRF. An examination of each screening process for differences in gender and Hispanic/White ethnicity will be conducted. The results obtained through these analyses may lead to better understanding of available screening tools and more accurate identification of students who need reading interventions in secondary schools in Colorado.

Research Questions

1. How does using the results of the Colorado State Assessment Program as a universal screening instrument to identify students in need of reading interventions compare with using the Test of Silent Contextual Reading Fluency?
2. How does using the results of the Colorado State Assessment Program as a universal screening instrument to identify students in need of reading interventions compare with using the Test of Silent Word Reading Fluency?
3. Are the results of the three screening tools different for males and females?

4. Are the results of the three screening tools different for students of different race/ethnicity: Hispanic/White?
5. Is there interaction of gender by Hispanic/White ethnicity for each of the three tests?
6. Do the three screening tools together add value to the process of identifying students in need of reading interventions?

Significance of the Study

This study may have significant impact on the implementation of universal screening in a response to intervention processes in secondary schools. Much attention has been given to RTI since the Individuals with Disabilities Education Improvement Act (IDEA) of 2004 (Public Law 108-446) permitted schools to use a student's response to scientific research-based intervention as part of process for identifying a specific learning disability (§1414 (B)(6)(A)). However, guidance for implementing RTI in secondary schools is limited. The use of appropriate and accurate universal screening tools will lead to identification of students in need of reading interventions and the implementation of interventions to address these needs. A universal screening process that is efficient, cost effective and accurate will prevent undue burden on staff, students and school infrastructures. Thus, assessment instruments that accomplish these goals, even at the secondary level, are vital if schools are to ensure academic success for all their students.

CHAPTER Two

Literature Review

This chapter provides a review of the prominent literature on Response to Intervention (RTI) and the essential components of the RTI model, with particular attention to universal screening at the secondary level. Key questions answered by this review are:

1. What is the typical profile of an adolescent struggling reader?
2. What is currently known about RTI and implementation at the secondary level?
3. What is known about the key RTI component universal screening?
4. What is known about the way middle and high schools implement universal screening in RTI frameworks?

Reading Profiles of Adolescent Struggling Readers

Despite the progress that has been made over the past 10 years in understanding and addressing the reading problems in elementary school students (McCardle & Chhabra, 2004), more than eight million adolescents continue to struggle with basic reading (Kamil, 2003). Reading involves a number of complex skills including word identification, phonemic awareness, comprehension, reading fluency and vocabulary (National Reading Panel, 2000). For adolescent students, reading comprehension becomes critical to learning content. Successful readers have mastered word level skills–

phonemic awareness, decoding, sight word reading and fluency, language comprehension skills–vocabulary, syntax and metacognitive strategies and use these to integrated text material with prior knowledge to derive meaning (Deshler & Hock, 2007). However, a significant number of adolescents struggle in reading.

A review of the literature indicates significant numbers of studies examining the profiles of young struggling readers exists, however fewer studies focus on adolescent struggling readers. A study by Buly and Valencia (2003) examined the reading skills of 108 fifth grade students who had performed below proficiency on the Washington State assessment, Washington Assessment of Student Learning (WASL). The purpose of their study was to define the specific reading deficiencies exhibited by students who had failed the reading portion of the state test. The study sample included 43% minority students. Results indicated students who performed poorly on the state assessment also performed poorly on all measures of reading: word identification, phonemic awareness, vocabulary, fluency and comprehension. Further analysis indicated three factors—word identification, fluency and meaning—accounted for 78% of the variance on the WASL scores. The study categorized the students as follows:

- Word callers (33%) with stronger word identification and fluency skills than vocabulary meaning, of which 15% had some difficulty with word identification
- Word stumblers (18%) with relatively strong meaning, but difficulty with word identification
- Slow and steady comprehenders (24%) with weak fluency but relatively strong word identification and comprehension

- Slow word callers (17%) with poor word level fluency and stronger word identification
- Disabled readers (9%) with poor skills in all areas.

The researchers concluded that poor performance on the state assessment was primarily due to poor reading fluency and comprehension (Buly & Valencia, 2003).

Leach, Scarborough, and Rescoria (2003) examined the cognitive and achievement profiles of adolescent students identified as reading disabled and compared them to those of early-identified disabled readers. The sample comprised of 161 students in fourth and fifth grades, of which only 5% were minorities. This study categorized the students as follows:

- No reading difficulties (59%)
- Reading comprehension difficulties only (8%)
- Word level difficulties only (17%)
- Both reading comprehension and word level difficulties (16%)

For those students identified as having some form of reading disability (41%), 35 % had adequate comprehension but word processing weaknesses; 32% had adequate word processing skilled but poor comprehension; and 32% had weaknesses in both. Thus, these struggling adolescent readers demonstrated weaknesses in both word level skills and comprehension,

A longitudinal study by Catts, Hogan and Adlof (2005) examined the language and reading abilities of a large group of students through second, fourth, and eighth grades. They categorized the 527 students identified in their study as reading disabled into one of three groups: students with adequate comprehension skills, but poor word

recognition; students with adequate word recognition, but poor comprehension; and students with weaknesses in both. The researchers found that 66% of eighth grade students had reading deficits related to poor comprehension, whereas 49% had weak word recognition.

A more recent descriptive study was conducted to examine the component reading skills of adolescent struggling readers attending urban schools (Hock et al., 2009). The sample was comprised of 345 late eighth grade and early ninth grade students, 202 of which were identified as struggling readers based on their performance on the Kansas state assessment, Kansas Reading Assessment. They used a cut-off point at the 40th percentile because "students performing at this mark are almost one-third of a standard deviation below the expected mean standard score, and thus below the expectation set by NCLB that all children read at grade level" (Hock et al., 2009, p. 25). Racial and ethnic representation included 52% African American, 15% Hispanic and 29% White students. In the final analysis, 202 adolescents were identified as struggling readers. Results indicated that these struggling readers performed one standard deviation below the mean in each of the reading areas (word identification and decoding, fluency, vocabulary and comprehension) and 20 or more standard score points lower than those students identified as proficient readers. In addition, 12% of the struggling readers had weaknesses in fluency and comprehension, but adequate word identification and decoding; and 61% had word level (identification and decoding) and comprehension weaknesses.

In summary, it appears that adolescent struggling readers can display both word level and comprehension skill deficits, with more recent research indicating that more

adolescent students have weaknesses at the word level than may have previously been demonstrated by earlier studies.

Response to Intervention

RTI is a multi-tiered system of prevention and intervention to address the needs of all students as they learn the content defined by the state standards and benchmarks. Figure 1 illustrates the RTI process: universal screening to identify students who are at risk for not achieving adequately, interventions those students identified at-risk, student progress monitoring, and intensified interventions for those students who continue to demonstrate poor or very limited response to the intervention.

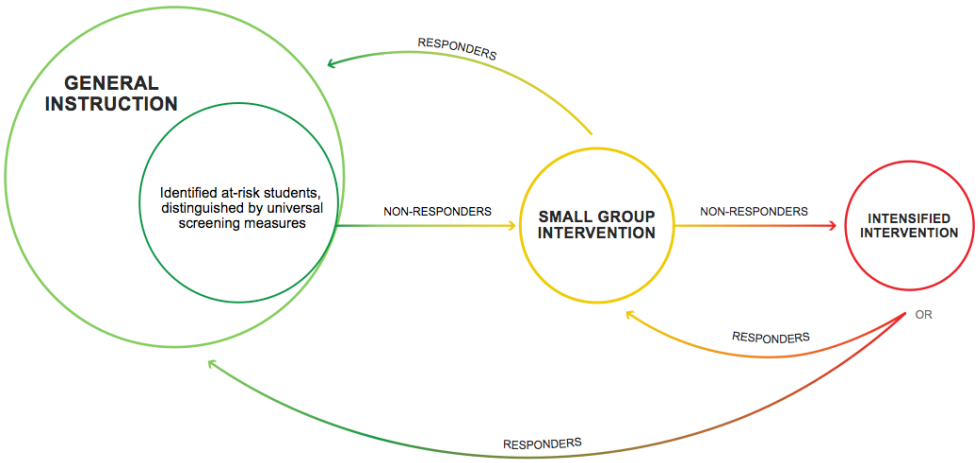


Figure 1. Generalized Model of RTI. All students receive universal screening, which distinguishes students at risk from those who are not. Students who are not responding to general instruction are placed in smaller groups for higher intensity instruction. Their progress is monitored, and those students responding to this intervention return to general education instruction. Students not responding to small group instruction are given more intensive instruction, often in a one-on-one setting. RTI models address responders at this more intensive level in different ways, including returning them to small group or general education instruction. The placement and provision of special education services within this framework varies by model.

The concept of prevention and intervention has had a presence in education for more than 20 years. The School Health Policies and Programs Study (Collins et al., 1995) collected information from all states and the District of Columbia, a nationally representative sample of public and private districts as well as middle/junior high and senior high schools, and a sample of randomly selected health education teachers. Results indicated that some instruction in prevention, as part of health and mental services education, was provided in schools at all levels in all states. Durlak (1997) summarized effective prevention programs for social and behavioral problems in children and adolescents that produced "significant and meaningful reductions in future problems, improvements in adaptive functioning, or both, and gains that were maintained are often maintained over time" (p. 23). In addition, over 90% of all school districts in the nation received federal funding to help children identified as at risk (Slavin, Karweit, & Madden, 1989).

Deno and Mirkin (1977) and Bergan (1977) authored two of the earliest studies using an intervention framework based on student response to instruction for academics and /or behavior. Deno and Mirkin used measurement based on curriculum to determine students' reading skills. The researchers describe a tiered system of services and resources for students whose performance was not adequate in relation to expectations of performance of peers in the classroom. They referred to this approach as the Data–Based Program Modification framework, which is an early version of what is now called a standard treatment protocol model of RTI.

Bergan (1977) conceived of a consultation model in which a four-stage process is applied to identify behavioral problems: measurement of behaviors discrepant from

desired or expected behavior, analysis of the problem to identify variables that may facilitate a solution, implementation of a plan for addressing the problem, and evaluation of the plan by examining the student's response (Bergan & Kratochwill, 1990).

Mellard and Johnson (2007) describe RTI models as containing common key components: universal screening, progress monitoring, levels or tiers of intervention, and fidelity of implementation. These components together provide a system of high quality research-based instruction, monitoring of each student's performance as learning occurs, provision of preventions to those students at risk, provision of interventions to those students not responding, and measurement of fidelity of implementation of the intervention.

Johnson, Mellard, and McKnight (2006) indicate the purpose of universal screening is to distinguish between those students who have significant academic or behavior weaknesses or are at risk of developing them from those who do not. Universal screening utilizes an efficient, low-cost, repeatable test of age-appropriate essential skills to identify students in need of interventions. Screening is usually conducted two to three times per school year. Pre-determined decision criteria or "cut points" are used to determine the students in need of intervention.

Johnson and colleagues (2006) define progress monitoring as the process of assessing student progress in the general education instruction or intervention program and as similar to what many educators call formative assessments. Staff and school teams use data obtained through progress monitoring to make decisions regarding students' response to the curriculum, instruction, or intervention and determine changes that need to be made in order to improve students' response rates. Progress monitoring tools should

be sensitive to small increments of growth over time, be administered frequently, are relevant to instruction and curriculum, and should result in data summarized in a clear concise manner that measures student progress and compares student performance with peers (Johnson et al., 2006). One method of progress monitoring is through Curriculum-based Measurement (CBM; Mellard & Johnson, 2007). CBM systematically assesses, as frequently as once a week, the different skills addressed by the curriculum or intervention.

Interventions are typically implemented to provide explicit instruction to meet the needs of students who are at risk or non-responsive to instruction or intervention (Johnson et al., 2006). Interventions are frequently viewed in a three-tiered model, with Tier 1 as the primary level of prevention in the general education classroom. Approximately 80% of a school's students should demonstrate adequate progress through Tier 1 instruction that is systematic, direct, explicit and based on scientific-research based curricula and practices. Tier 2 supplements the curriculum and general education instruction by providing interventions to students who fail to progress when provided high quality general education instruction and strategies. Tier 3 is the most intense level of intervention and in some cases is synonymous with special education. This level of intervention is intensified in duration and frequency with weekly or even daily progress monitoring to improve the learning of those students who to do not respond or respond poorly to the interventions in Tier 1 and Tier 2.

Fidelity of implementation is an assessment of whether instruction was delivered in the manner it was intended (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). Fidelity is important both at the building and teacher level, and should be

implemented for the RTI model as a whole in addition to the individual components of RTI.

A number of studies have been conducted on implementation of RTI, mostly at the elementary level. A quasi-experimental study with elements of historical contrast was conducted of the St. Croix River education district model, a multi-tiered problem-solving model, to determine its impact on student achievement in reading (Bollman, Silberglitt, & Gibbons (2007). Data were analyzed from five districts implementing the model from 1995 to 1996. Bollman and colleagues reported overall improvement in student reading scores in this study. Callender (2007) conducted a quasi-experimental descriptive study of the Idaho Results Based Model as part of a five-year evaluation process. They concluded students with an intervention plan through the model demonstrated more progress in reading than their counterparts not in the model. The Minneapolis problem-solving model was studied using an historic contrast design; the authors reported a decrease in the special education identification rate (Martson, Muyskens, Lau, & Canter, 2003). A descriptive study was completed on the Illinois Flexible Service Delivery Model using data from 26 schools across the state (Peterson, Prasse, Shinn, & Swerdlik, 2007). The authors reported student improvement in both behavior and academics. In addition, a meta-analytic review of RTI research examining field-based and research-implemented models of RTI computed effect sizes and unbiased estimates of effects for 24 studies (Burns, Appleton, & Stehouwer, 2005). Results indicated unbiased estimate of effect exceeded 1.0. Although more studies are need on RTI models, these results, in general, are promising.

A number of studies have been conducted on components of the RTI model at the elementary level, particularly in the area of reading interventions. Vaughn, Linan-Thompson, and Hickman (2003) conducted a study of 45 second-grade students identified as being at risk for learning disabilities based on a screening measure. Students received supplemental instruction for 10 weeks followed by an assessment to determine if progress was adequate to exit this instruction. Supplemental instruction continued into a Tier 3 and Tier 4 level, as needed. Results indicated 34 of the 45 students progressed adequately to exit some level of supplemental instruction. Tiers of intervention for students in kindergarten through third grade were the subject of an historical contrast design study by O'Connor, Harty, and Fulmer (2005). Students in these grades (100 students per grade) received evidence-based instruction for three months. Those non-responsive students received small-group instruction as a Tier 2 intervention. Those students who continued to be non-responsive received Tier 3 intervention in the form of one-to-one instruction. Students who received tiered instruction demonstrated improvement on all reading measures when compared to the historical contrast group.

A number of other studies with foci on reading interventions in the elementary grades have been conducted, and results indicated more progress for those students who received interventions for specific reading deficits compared to students who did not participate in the specific interventions (Foorman, Winikates, Mehta, Schatschneider, & Fletcher, 1997; Vaughn et al., 2003; O'Connor et al., 2005; Fuchs et al., 2005; Fuchs, Fuchs, & Compton, 2004). Two studies, utilizing quasi-experimental design, examined Hispanic learners and their responses to reading interventions (Gunn, Smolkowski, & Ari, 2000; Gunn, Smolkowski, Biglan, & Black, 2002). Results indicated Hispanic

students demonstrated significant gains compared to non-Hispanic students after receiving reading interventions that addressed specific reading skills.

Studies involving math interventions are limited. Ardoin, Witt, Connell, and Koenig (2005) conducted a single case, A-B design study of 14 fourth grade students in an elementary school. Screening and a class wide implementation took place. Data indicated that nine of the students improved over baseline, and peer tutoring assisted five other students.

The studies referenced above each included some form of universal screening to determine students in need of intervention. The screening tools used went beyond the summative data obtained from state assessments to identify students at risk or demonstrating skill deficits in a particular academic area. Ardoin et al. (2005) and Fuchs et al. (2005) used CBM math probes to determine students at risk for below standard end of year performance. Other studies used such reading screening tools as the Early Reading Screening (Morris, Tyner, & Perney, 2000; Santa & Hoien, 1999), Dynamic Indicators of Early Basic Literacy Skills or DIBELS (Gunn et al., 2000; Gunn et al., 2002), or such norm-referenced tests as the reading subtest of the Wide Range Achievement Test-Revised (Jenkins, Peyton, Sanders, & Vadasy, 2004).

Response to Intervention at the Secondary Level

Although research at the elementary level has been considerable, no comprehensive research studies on the implementation of the RTI model in secondary schools exist, and only a few studies and papers have been presented on one or more components of RTI at the secondary level. To examine progress monitoring, Epsin, Scierka, and Skare (1999) tested 147 students in tenth grade using CBM in written

expression. They compared various assessments measuring students' writing abilities (including number of words written, words spelled correctly, characters per words and sentences written). They assessed the effect of progress monitoring on student achievement through students' general writing proficiency, including California state achievement test scores, first- and second-semester English grades, and considered differences by students' group placement (SLD, basic English, regular English, advanced English). Results indicated secondary level CBM procedures for writing need to be more complex and require a larger time commitment and more instruction by teachers than those at the elementary level.

A second study of CBM examined the effects of peer-assisted learning strategies (PALS) and CBM on math performance in a high school (ninth through twelfth grade) setting (Calhoun & Fuchs, 2003). This study implemented PALS twice weekly with once weekly CBM for 15 weeks. Findings from this study indicated the combination of PALS and CBM resulted in improved computation math skills, however skills in math concepts and application were not statistically different from pre-test measures.

Twyman and Tindal (2007) and Ketterlin-Geller, McCoy, Twyman, and Tindal (2006) investigated the reliability of a concept maze task to assist middle school teachers in making accurate decisions regarding students' content learning. A concept maze is a task that requires students to select the best answer from a list of possible choices to complete a sentence, thus measuring content comprehension rather than overall general reading comprehension. Results support the use of concept mazes that focus on attributes as a measure of change in content areas. The researchers explained that a concept maze measure could be used in combination with other measures in a content area to determine

to what extent students are on par with or discrepant from their peers at a single point in time, as well as how they are progressing individually and relative to classmates.

In the area of interventions, Mercer, Campbell, Miller, Mercer, and Lane (2000) examined the effects of a reading fluency intervention program used to supplement reading instruction for middle school students identified as having Learning Disabilities. The intervention focused on specific reading skills (phonics, sight word phrases, and oral reading fluency) with varying duration (ranging from 6 to 25 months). Students demonstrated significant growth in reading level and fluency. A second study focused on reading interventions for students repeating eighth grade in a large urban school setting. Students who received the intervention showed significant gains (over three normal curve equivalents) in reading (Papalewis, 2004). In contrast, mixed results emerged in another study involving implementation of a small-group reading intervention with struggling readers in one middle school for one semester. After assessment in specific reading areas, students received instruction five times per week, 50 minutes per session, for 12 weeks. Some students demonstrated progress, whereas others made little or no progress in either the intervention or traditional reading class (Denton & Wexler, 2006).

Several case studies (Johnson & Smith, 2008; Bacon, 2005; Duffy, 2007; Fisher, 2001) have examined RTI in secondary schools. Even as case studies do not meet the rigor of scientific research as defined by No Child Left Behind, they do have an important place in the literature. Case studies form the foundation for understanding and formulating future investigations, suggest hypothetical directions for the use of more rigorous research studies that meet the scientific definition, and provide important additional data. Johnson and Smith (2008) described the implementation of RTI at a

junior high school during the 2006-07 school year. Their case study found that the existing school-wide behavior support system served as the foundation for implementing RTI in academic areas. At the end of the year, they reported that the systemic process of evaluating student performance and the focused efforts on improved instruction and interventions resulted in reduced referrals for special education evaluation. Another case study described one teacher's efforts to provide specific strategy skills through a coaching model to students who struggled in reading. Bacon (2005) noted that as a result of the intervention, the students attained a higher instructional reading level. The case of the Long Beach Unified School District referred to by Duffy (2007) employed regular assessments and tiered interventions as best practice in meeting the needs of all students. They provided students entering high school two or more years behind in reading additional literacy instruction as well as core literacy instruction, including a double block of language arts or an after-school reading program. Duffy reported that this district had such success that they moved implementation of this system to the middle school level. A fourth case study described the RTI efforts of an urban school to address students' literacy weaknesses. Fisher (2001) highlighted two components beyond those related to in RTI that need to be addressed at the secondary level: professional development, and scheduling time for effective daily reading instruction. Although these case studies lack the scientific evidence one expects in research, they do indicate the merits of implementing an RTI system that provides interventions to struggling students.

Universal Screening

"Screening is a type of assessment that is characterized by providing quick, low-cost, repeatable testing of age-appropriate critical skills" (Johnson et al., 2006, p. 1.2). The purpose of screening is to identify those students who are at risk for failure in acquiring skills defined by state standards and benchmarks, despite receiving effective scientific, research-based instruction. In RTI, screening is conducted for all students and is typically conducted three times a year (Hughes & Douglas, 2009).

Screening, as a strategy for early identification of people in need of attention or prevention for diseases and health problems, is used on a large scale in the field of medicine. Such screening is a two-stage process in which a large group is assessed using brief, low-cost tests to sort out individuals who have a health problem or are at risk of developing one and thus need further diagnostic tests. One example is that of tuberculosis screening. Rather than giving all people a chest x-ray, which can be costly and time consuming, a prick test is used and only individuals who test positive on this screening tool are given a chest x-ray (Lichtenstein & Ireton, 1984).

Application of screening concepts to an educational setting seems feasible and may be at the core of processes that ensure success for all students. However, the key to implementing universal screening is to identify the best screening process for the specific situation. Lichtenstein and Ireton (1984) proposed a screening framework that included "(1) clarifying the needs and constraints of the particular screening situation, (2) gathering information on prospective measures, and (3) evaluating alternative choices on the basis of meaningful selection criteria" (p. 102). In the context of RTI, the process of

clarifying needs and constraints of the screening situation includes determining if the need is identifying weaknesses in basic skills of reading, math, and writing applied across content areas or in content-specific knowledge areas. Resources related to the screening process that need to be examined include the cost of materials; time required to administer the measure; and staff with expertise required for administering, scoring, and interpreting the measure and results. Gathering information on prospective measures includes examining the sensitivity and specificity, reliability, consequential validity and, to a lesser degree, how the instrument integrates within the intervention model (Jenkins, 2003).

Sensitivity and specificity deal with the accuracy of the screening measure. The concept's origin can be traced to statistical decision theory first introduced by Abraham Wald (1950), which in the broadest terms deals with "the problem of decision making in the face of uncertainty" (Girshick, 1954, p. 448). Decision-making theory applied to education provides a means to describing complex problems, identifying a course of action, assessing the probability of the outcomes, and calculating the most effective decision (Harber, 1981). This theory was applied to formulate the often written about "hit rate model" for analyzing outcomes (Barnes, 1982; Frankenberg, 1974; Meisels, 1985; Wilson & Jungner, 1968; Lichenstein & Ireton, 1984). The model summarizes the relationship between screening outcomes and actual status of an individual. Figure 2 illustrates the hit rate model.

		"Actual" Status: Child needs special services	"Actual" Status: Child does not need special services
Screening	Refer:	A accurate referral; Valid positive	C over referral; false positive
	Do not	B under referral; false negative	D accurate non-referral; valid negative

Problem base rate: $\frac{A + B}{A + B + C + D}$

Referral rate: $\frac{A + C}{A + B + C + D}$

Sensitivity: $\frac{A}{A + B}$

Specificity: $\frac{D}{C + D}$

Efficiency of screening outcome "refer": $\frac{A}{A + C}$

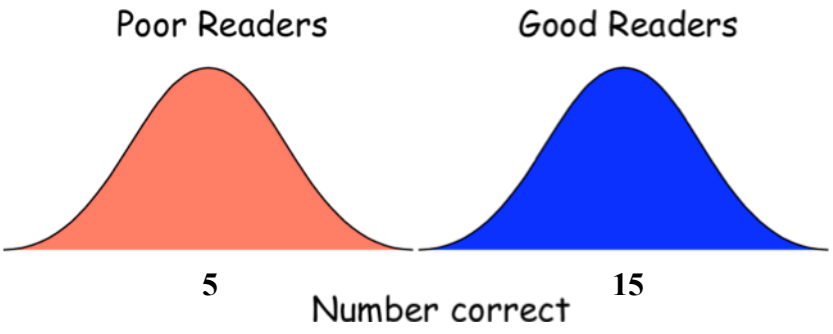
Efficiency of screening outcome "do not refer": $\frac{D}{B + D}$

Figure 2. Hit Rate Model. "Possible outcomes of screening are that a child is either categorized as a screening positive, meaning the child is regarded as high risk and referred for further assessment; or as a screening negative, meaning the child is low risk and not referred" (Lichtenstein, & Ireton, 1984, p. 197-198).

Sensitivity of a screening measure is its accuracy of identifying students at risk (true positives), whereas specificity is its accuracy of identifying those students not at risk (true negative). Johnson et al. (2006, p. 1.2) suggest, "it is better to err on the side of false positives (students identified as at risk, who through more intense assessment are found to have been misidentified) than on the side of false negative (students not identified through screening who later turn out to be at risk)". This screening approach is especially relevant at the secondary level, when time to address problems is limited by the number of years a student has left before exiting the K-12 education system.

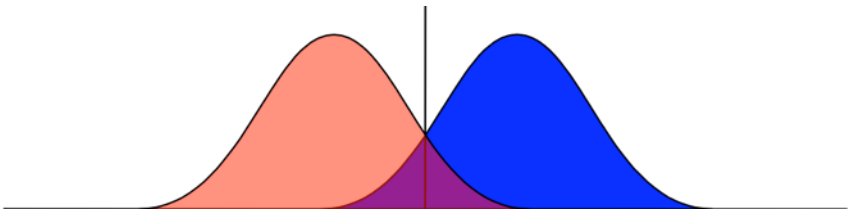
Accuracy is also obtained through the decision point or cut score that represents the dividing line between those students at risk and those not at risk. The cut scores should be determined as part of the screening process and applied consistently. Ideally decision rules would result in 100% true positives and true negatives; however, because screening is intended to determine risk and not diagnose or determine eligibility of a disability, 80 % true positives and true negatives may suffice (Catts, 2006). Figure 3 illustrates this concept.

Ideal Screening



TP 100	FP 0
FN 0	TN 100

Typical Screening



TP 80	FP 20
FN 20	TN 80

Figure 3. Distribution from Cut Scores. The ideal screening would result in sensitivity and specificity of 100%. Since screening is not used for diagnosis, a sensitivity and specificity rate of 80% is adequate (Catts, 2006, p. 21-22).

Reliability is the degree to which an assessment or test yields similar results each time it is used under the same condition with the same subjects. A good screening measure is one that is reliable enough to yield consistent similar results thus giving one confidence that the results are close to true (Lichtenstein & Ireton, 1984). Variability in the student, the situation, or the examiner could all influence the screening reliability. Test-retest, equivalent forms, internal consistency or inter-rater reliability can all be used to determine a screening measure's degree of reliability. Test-retest reliability is the comparison between results obtained on two separate testing sessions and is considered a more conservative method of estimating reliability. One of the concerns with this method is deciding how much time should elapse between the administrations. If the interval is too short, scores could be influenced by test recall; if the interval is too long, scores could be influenced by changes in the subjects. Generally, as the length of time increases from one session to the next, the correlation between the scores decreases (Johnson & Christensen, 2004). A correlation of .90 is recommended for diagnostic tools while a correlation of .80 is acceptable for screening measures (Lichtenstein & Ireton, 1984).

Validity is the degree to which inferences about the subjects obtaining particular scores on a test can be made. Validity questions fall into two categories, inferences about what is being measured by the test and those regarding the usefulness of the test (Messick, 1980). Validity has been addressed in depth and as far back as the 1940s (Jenkins, 1946; Gulliksen, 1950; Cronbach & Meehl, 1955; Guion, 1974). The first category—does the test measure the characteristics it is interpreted to test—can be answered by examining the test's psychometric properties provided by the test publisher.

The second category deals with what Messick (1980) refers to as consequential validity—does the screening tool avoid inequitable treatment, consume resources efficiently and effectively and link to effective interventions (Jenkins, 2003). When screening to identify students at risk in reading, a tool needs to be suitable for measuring the constructs of reading and determining the performance level that will distinguish those at risk from those not at risk of academic failure. The tests developed by Woodcock and associates, including the Woodcock-Johnson III Tests of Achievement (Woodcock, McGrew, & Mather, 2001) and the Woodcock Reading Mastery Test-Revised (Woodcock, 1998) are considered the "gold standard" of reading assessments (Jenkins, 2003). However, these tests are individually administered by a professional with extensive training and take a considerable amount of time to complete, thus being inefficient and cost prohibitive to use as a screening tool.

Jenkins' (2003) summary of the research on screening measures used in kindergarten through second grade included sensitivity and specificity percentages. Each screening measures included some aspect of phonological awareness and letter identification in kindergarten and first grade. Multiple assessments used in combination resulted in higher specificity and sensitivity. Currently many elementary schools use tools like the DIBELS (Good, Simmons, & Kame'enui, 2001) or the Texas Primary Reading Inventory (Foorman et al., 1998). Few studies of screening measures have been conducted beyond second grade.

No studies regarding screening measures in RTI models at the secondary level exist. Studies that have included CBM are extant; however, these studies use CBM as a means of progress monitoring rather than screening (Epsin & Halverson, 1999; Calhoun

& Fuchs, 2003; Twyman & Tindal, 2007; Ketterlin-Geller et al., 2006). The authors of the major models of RTI agree that screening is an important part of the process, however the features of screening vary in the tool, method, and frequency (Batsche et al., 2008; Chun & Witt, 2008; Fuchs & Fuchs, 2005; Johnson et al. 2006; Vaughn, 2005; Shinn, 2008; Sugai & Horner, 2007).

Batsche, Curtis, Dorman, Castillo, and Porter (2008) recommend early assessment of all students, especially kindergarteners, within the first 30 days of school in order to timely deliver interventions. During the school year, screening data is used to determine if the core curriculum is effective (80% of students making benchmarks), identify which students are at-risk for failure, and assess if over-representation of certain student groups exists in the at-risk population. The basic components of RTI are applied first to all students in a school to determine what percentage of the students will respond to the core curriculum. Six areas need to be considered to ensure the problem-solving process is ecologically sound and culturally competent: student factors, teacher factors, peer factors, curriculum factors, classroom/school factors, and family/community factors.

Chun and Witt (2008) utilize brief CBM measures of reading fluency, reading comprehension, math, math application, early numeracy and literacy, as well as writing. The reading fluency probes were written by sampling from a large word frequency database of medium to high frequency words for each grade level and were subsequently evaluated validity, reliability and predictive accuracy of the probes and decision rules have been evaluated (Witt & VanDerHeyden, 2007). Results of a study by Ardoin, Witt, Suldo, Koenig, Resetar, Slider, and Williams (2004) indicated that one CBM probe yields equivalent results to three CBM probes.

The “maze” universal screening process has been evaluated using various curriculum-based as well as normative assessments as the criterion variable (Van Hook, 2008). Validation studies indicate criterion and concurrent validity as well as classification accuracy are equivalent or greater than oral reading fluency probes for students in third grade and above. Screening data are used to determine if core instruction is effective and to detect the presence of issues such as disproportionality or differences between classes. Screening data are also used to examine individual, class wide and school wide student performance level and progress during the school year. Decision rules for screening have been evaluated and validated in peer-reviewed studies examining concurrent and predictive validity as well as classification accuracy (e.g., ability to identify true positives and to rule out true negatives). These studies provide direct evidence of validity from both a classical measurement as well as a classification accuracy perspective (Witt & VanDerHeyden, 2007). Administration of the screening probes follows a scripted implementation protocol, including written instructions and the use of a countdown timer (VanDerHeyden & Gilbertson, 2008).

Fuchs and Fuchs (2005) recommend that typically in the first month of school, a criterion to indicate at-risk status is determined, such as scores below the 25th percentile on the previous year's high stakes test. Schools test all students, and consider at risk those students scoring below the criterion percentile or performance benchmark. The Fuchs further recommend administering a screening tool with a benchmark that can be used for predicting end-of-year performance on high stakes tests or local graduation requirements. To reduce false positives, students identified by the screening tool as being at-risk should be progress monitored for five to eight weeks. Those students who do not demonstrate

adequate progress when compared to peers or who fail to demonstrate adequate individual growth can be considered non-responsive and receive a Tier 2 level of intervention (Fuchs & Fuchs, 2005; Fuchs & Deshler, 2007).

Johnson, Mellard, and McKnight (2006) concur with the Fuchs recommendations and also suggest screening all students three times per year to identify any additional students who are not responding to the instruction throughout the year. They posit that screening data be used not only for screening purposes, but also to determine if students are progressing at an acceptable rate (benefiting from intervention at Tier 1), or if they need a change in the intervention program (move to Tier 2 or Tier 3). Data from screening can also assist in curricular and instructional decisions.

Sugai and Horner (2007) focused on behavior and prescribed widespread data collection on factors, including attendance and office referrals, and comparisons at school, class, and student levels to support balanced decision making. For example, to determine who are responsive to existing behavior practices and systems at the school-wide discipline level, increasing attention may be focused on students who have 0-1, 2-5, or 6 or more major rule violations.

Two authors focus on screening at the secondary level. Shinn (2008) suggests the use of screening with CBM maze reading tasks with 5th, 6th and 9th grade students. He also recommends that individualized, targeted screening be the focus at the high school level rather than universally screening all students. Vaughn (2005) recommends screening in word identification, comprehension, fluency and spelling for middle school.

Summary

RTI developed from research dating back to the early 1990's. Its application in education began in the school health area and evolved to be used to address both academic concerns of poor performing students. Studies indicate that RTI at the elementary level is occurring with some promising results in improved student achievement and decreased number of students referred and found eligible for special education services. Most research studies have been conducted at the elementary level, and studies on specific components of RTI and case studies have been conducted in recent years.

Universal screening is a key component to RTI that identifies students at risk for developing significant problems in specific academic areas. Considerable research exists regarding the essential characteristics of an effective screening tool, including high sensitivity and specificity, reliability, validity (both context and consequential) as well as efficiency and high cost effectiveness. Several high quality tools have been developed and implemented at the elementary level, particularly in the area of reading; however, the field lacks an effective screening measure at the secondary level.

Screening could be accomplished using results from state summative or norm-referenced assessments (Jenkins, 2003). Some schools use the results of state assessments for identifying students at risk. One study used the state assessment, CSAP, as a school wide screening tool (Johnson & Smith, 2008). A recent study presented at the Council for Exceptional Children Annual Conference used the results of the California state assessment to identify students needing intervention in reading in sixth grade (Graves, Block, & Brandon, 2009). Schools may use of state assessment results to make screening

decisions due to a lack of availability of efficient, cost effective screening tools at the secondary level or the ease, and availability of the state results. CBM is also used for screening, but is primarily used for progress monitoring. Some researchers recommend following screening with progress monitoring, using CBM to ensure accurate identification of those students in need of intervention. In addition, researchers have different views regarding screening tools at the secondary level, particularly about whether to use locally-normed tools, criterion-referenced tools or nationally norm-referenced tools. Finally schools need to determine decision rules prior to implementing a screening process, and those rules should be sensitive enough to accurately identify students at risk with the least amount of false negatives.

A significant number of middle and high school students continue to struggle with reading at the word level, comprehension level, or both. Accurate assessments are vital to complete in order to identify the nature and severity of the reading issues so that interventions can be matched to need. Universal screening tools need to be reliable, valid, cost-effective, efficient, and readily available for use at the secondary level. The need for more research is apparent, particularly at the secondary level, as RTI is being implemented to address the reading weaknesses of students in secondary schools.

Chapter Three

Methodology

A review of the literature illustrated that an essential component of Response to Intervention (RTI) is universal screening to identify students at risk or demonstrating poor academic skills such as reading. Once identified these students' schools can provide increasingly intensive interventions to improve skills and close the gap in performance from those students not at risk. For screening measures to be useful, they should identify students who require further assistance, be practical, and accurately identify students (Jenkins, 2003). Universal screening is characterized as an efficient, low-cost, repeatable test of age-appropriate critical skills (Johnson et al., 2006, p. 1.2) having an accuracy rate of 80% or greater (Catts, 2006; Jenkins, 2003; Lichtenstein & Ireton, 1984).

Research Design

This inquiry is a comparative study of three tools for universal screening. Students identified as needing reading interventions based on the proficiency levels of their scores on CSAP are compared with those students identified using the results from the TOSCRF (Hammill et al., 2006) and those identified using results from the TOSWRF (Mather et al., 2004). In addition each tool is examined for significance by gender and Hispanic/White ethnicity.

Specific research questions are:

1. How does using the results of the Colorado State Assessment Program as a universal screener to identify students in need of reading interventions compare with using the Test of Silent Contextual Reading Fluency?
2. How does using the results of the Colorado State Assessment Program as a universal screener to identify students in need of reading interventions compare with using the Test of Silent Word Reading Fluency?
3. Are the results of the three screening tools different for males and females?
4. Are the results of the three screening tools different for students of different race/ethnicity: Hispanic/White?
5. Is there interaction of gender by Hispanic/White ethnicity for each of the three tests?
6. Is value added by using the three screening tools together to identify students in need of reading interventions?

Sample and Sampling Procedures

The study uses the same sample of students throughout, and compares which of them are identified as at risk in the area of reading using the CSAP proficiency level results with at-risk identification using the results from the TOSCRF (Hammill et al., 2006) and the TOSWRF (Mather et al., 2004). The sample is from Colorado school districts because of the existence of data collected as part of a state reading project. The Colorado Department of Education (CDOE) initiated the Rural Secondary Literacy Project in the 2007-08 school year. The CDOE selected districts to participate in this project based on an application process. Only districts with a student enrollment of 2,000

or less, a rural district designation, a focus on secondary level literacy, and at least an 80% staff commitment to the project were eligible to apply. The CDOE uses the U.S. Census Bureau's (2006) definition of rural: territory, population and housing units not classified as urban that consist of a large central place and adjacent densely settled census blocks that together have a total population of at least 2,500 for urban clusters, or at least 50,000 for urbanized areas. CDOE selected 9 districts to participate in the project as either single districts or consortiums of three or more districts. All participating districts were required to implement a screening process that included administration of the TOSCRF (Hammill et al., 2006) and the TOSWRF Mather et al., 2004) in middle and high school grades.

Only the seven single districts in the state project were contacted to participate in this study. A letter (Attachment A) and follow up emails sent and phone calls were made to the districts requesting the following elements of their extant data for each student:

- Grade level
- Ethnicity
- Special education status
- ELL status
- Free and reduced lunch status
- CSAP 2007 score and proficiency level
- CSAP 2008 score and proficiency level
- Initial screening results
    - TOSCRF standard scores and percentiles
    - TOSWRF standard scores and percentiles

Three districts agreed to participate in this study, sharing data that already existed due to their participation in the state project, resulting in study data from a total of 1,763 cases. The districts consistently reported gender and ethnicity, however, they were inconsistent in reporting special education, ELL, and free and reduced lunch statuses. Therefore, analyses were conducted for gender and Hispanic/White ethnicity only. Cases without CSAP scores or proficiency levels were also excluded from analyses because the research questions required both these scores. After excluding cases with inconsistent and/or incomplete data, 649 cases in fifth through eleventh grade remained for analysis.

Procedures for the Protection of Human Subjects

The University of Denver Institutional Review Board procedures were followed, and approval for data collection was obtained. Student identifiable information was not collected. In addition, all references to district name and state identification number were excluded from the data collection and this report.

Instruments

This study uses three tests as screening tools for reading weaknesses. The first test, the Colorado State Assessment Program (CSAP), was chosen due to the frequency of its use by schools to identify students for targeted instruction. The reading portion of the CSAP is a summative reading comprehension test that assesses a student's ability to read and understand text. Students read, predict, summarize, comprehend, and analyze fictional and non-fictional texts; determine the main idea and locate relevant information; use word recognition skills; and make inferences, draw conclusions and respond to literature that represents different points of view of literature. The Test of Silent Contextual Reading Fluency (TOSCRF) (Hammill et al., 2006) also assesses reading

comprehension, however it is designed as a screening assessment rather than a summative assessment. The student uses skills in word identification, word meaning, word building, sentence structure, comprehension, and fluency to make meaning of what is read. The third screening tool, The Test of Silent Word Reading Fluency (TOSWRF) (Mather et al., 2004) is also a screening tool measuring overall reading ability using words from graded words lists for sight word recognition and vocabulary meaning. The TOSCRF and the TOSWRF both use word chains as the test stimuli however the TOSCRF improves upon the TOSWRF by embedding words within passages of text (Bell, McCallum, Kirk, & McCane-Bowling, 2007). While these tests differ in methods, they each measure overall reading ability, being able to understand what is read. More detailed descriptions of each test follow.

The purpose of the CSAP is to provide an annual measure of student performance relative to the Colorado Model Content Standards. The CSAP test is a timed, standardized instrument administered under standardized conditions as reported by the instrument designers, CTB/McGraw-Hill (2007, 2008). The reading, writing, and mathematics portions of the CSAP are administered to students in third through tenth grades; the science portion is administered to students in grades fifth, eighth and tenth grade students. The reading portion is based on the reading comprehension model content standard: students read and understand a variety of materials, and content areas of fiction, nonfiction, vocabulary, and poetry. CTB/McGraw Hill reported content related validity and reliability; however did not include actual statistical information in their technical reports. Scores obtained are categorized in one of four proficiency levels: Unsatisfactory,

Partially Proficient, Proficient and Advanced. This assessment is a summative assessment measuring overall reading comprehension as defined by the state standards.

The TOSCRF (Hammill et al., 2006) measures the silent general reading ability of students ranging in age from 7 years 0 months to 18 years 11 months. The authors define contextual reading as using skills in word identification, word meaning, word building, sentence structure, comprehension, and fluency to make meaning of what is read. The TOSCRF can be group administered in ten minutes by classroom teachers, reading specialists, or other staff with knowledge of standardized test administration. This test yields a raw score, percentile rankings, standard scores, and age and grade equivalents, and was normed using a national representative sample of 1,098 individuals residing in 23 states. The students in the norming sample represented a broad range of reading ability, ethnicity, and disabilities. The authors report from their studies reliability coefficients ranging from .81 to .95. Correlation coefficients between the TOSCRF and other reading measures range from .61 to .89. The authors report that the TOSCRF accurately and reliably identifies students who are struggling with reading, and thus can be used as a screening method to determine reading difficulties, specific needs for remediation, and types of intervention required. The results of an independent study examining the psychometric integrity of the TOSCRF supported the use of it as "a quick screener and as a gross measure of reading progress" (Bell, McCallum, Kirk, & McCane-Bowling, 2007, p. 46).

The TOSWRF (Mather et al., 2004) measures word identification, fluency, and word comprehension; thus it is considered a valid estimate of general reading ability and can be used to identify poor readers. TOSWRF can be group administered to both

elementary and secondary students in three minutes by a classroom teacher, reading teacher, or other staff member with minimal training in assessment. The assessment has two equivalent forms making it suitable for screening purposes. The instrument was normed using a representative sample of more than 3,592 individuals ranging in age from 6 years 6 month to 17 years 11 months, who represented a variety of demographic subgroups including ethnicity and disabilities, and who lived in 34 states. The TOSWRF yields a raw score, standard scores, percentile rankings, and age and grade equivalents. Reliability was examined using four types of reliability: alternate form-immediate administration, test-retest, alternate form-delayed administration, and scorer differences. Coefficients ranged from .86 for alternate form-immediate administration to .99 for inter-rater reliability. Correlations of the TOSWRF with other tests measuring word identification and fluency, like the Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999), were greater than .70. Specificity and sensitivity rates ranged from 84% to 89% in five of the six studies, which is considered adequate for screening purposes. Results of an independent study of the concurrent validity of the TOSWRF support its use, "particularly for screening for relatively weak readers" (Bell, McCallum, Burton, Gray, Windingstad, & Moore, 2009, p. 8).

Decision Rules for the Screening

RTI requires a screening process to identify students in need of interventions. Decision rules need to be determined prior to beginning the screening process and were defined by the researcher of this study. When the CSAP is used as a screening tool, students whose proficiency level on the reading portion is Unsatisfactory or Partially Proficient are considered at risk and thus in need of interventions. Tier 2 interventions

would be provided to students receiving a Partially Proficient performance level, whereas those students with an Unsatisfactory performance level receive Tier 3 intervention.

The decision rule for the TOSCRF and TOSWRF in this study is based on percentile scores. A student whose percentile score is equal to or below the 16th percentile would be identified for a Tier 3 intervention. A student whose percentile score is above the 16th percentile but below the 40th percentile would be identified for a Tier 2 intervention. Any student whose percentile score was at or above 40th percentile was considered not at-risk. These decision points are consistent with the recommendations of researchers of RTI and adolescent struggling readers as noted in the literature review. In addition, these rules are consistent with expectations set by No Child Left Behind (Public Law 108-446), which requires all students read at grade level by the year 2014. Students whose reading performance is below the 40th percentile are almost one-third of a standard deviation below the mean, thereby reading below expected grade level.

Screening Procedures

Each of the three districts administered the CSAP in March and April 2007. The TOSCRF and TOSWRF were administered in fall of 2007. District-identified staff received training on the TOSCRF and TOSWRF screening instruments as part of their participation in the state project. A coach, assigned to each school as part of the project, reviewed the screening process and scoring. Each district identified a staff member to compile and keep the results. Each district used the results to determine the level of intervention needed for each student.

Data Collection and Analysis

Responding districts sent their data in electronic format in an Excel spreadsheet to the researcher. The spreadsheets did not contain any student identifiable data. Cases with missing data were excluded and a final data file of 649 cases was exported to statistical software for analysis. Each case contained grade level, gender, ethnicity, 2007 CSAP score and proficiency level, Fall 2007 TOSCRF percentile and standard scores, and Fall 2007 TOSWRF percentile and standard scores. The Friedman Two-Way Analysis of Variance for related samples, Chi-Square Test of Independence and Hierarchical Log Linear tests were completed for the analysis by gender and ethnicity, and the Wilcoxon Signed Ranks Test was used to address research questions one and two.

Chapter Four

Results

The purpose of this study is to examine three universal screening methods used to identify students in need of reading interventions at the secondary school level. One screening method, using the CSAP (CTB/McGraw Hill, 2007, 2008) proficiency levels, was compared to decisions made based on the TOSCRF (Hammill et al., 2006) and the TOSWRF (Mather et al., 2004) scores. In addition, difference in results by gender and ethnicity are reported.

Demographics

The three Colorado districts from which the data was drawn were rural—consisting of population and housing units not classified as urban which consists of a large central place and adjacent densely settled census blocks that together have a total population of at least 2,500 for urban clusters, or at least 50,000 for urbanized areas (U.S. Census Bureau, 2006)—and had student enrollments below 2,000. The ethnicity of the student population of the three districts averaged 1.1% American Indian or Alaskan native, 0.5% Asian or Pacific Islander, 1.2% Black, 25.5% Hispanic, and 71.1% White, non-Hispanic. Only one of the three districts made adequate yearly progress in reading for the year the data were collected. An average of 38% of the combined district student enrollment was economically disadvantaged.

After excluding students who did not meet the criteria of having CSAP scores and proficiency levels, TOSCRF and TOSWRF standard scores and percentile rankings, and

gender and ethnicity designations, a total of 649 students in fifth through eleventh grades remained in the sample dataset. Descriptive demographics statistics for the sample are presented in Tables 1 and 2. Students were distributed across the grades fairly evenly, with the exception of fifth grade with 62 students (9.6%) and sixth grade with 137 students (21.1%). The sample was evenly distributed by gender. However, the sample size did not include a large enough representation of American Indian or Alaskan Native, Asian, and Black (non-Hispanic) students to include these groups in analyses involving ethnicity. In addition, the Hispanic representation in the sample size is lower than in the three districts' enrollment, therefore limiting generalizability of study results to a larger population.

Table 1

Description of Study Sample

	n	Frequencies
Gender		
Male	360	55.5%
Female	289	44.5%
Ethnicity		
American Indian or Alaskan Native	2	0.3%
Asian	0	0.0%
Black (non-Hispanic)	19	2.9%
Hispanic	109	16.8%
White (non-Hispanic)	519	80.0%
Grade		
Grade 5	62	9.6%
Grade 6	137	21.1%
Grade 7	89	13.7%
Grade 8	107	16.5%
Grade 9	92	14.2%
Grade 10	88	13.6%
Grade 11	74	11.4%

Table 2

Comparison of Sample to Combined Districts Demographics

Ethnicity	Frequencies			
	Sample		Average of the 3 Districts	
	n	%	*n*	%
American Indian or Alaskan Native	2	0.3%	48	1.1%
Asian	0	0.0%	21	0.5%
Black (non-Hispanic)	19	2.9%	53	1.2%
Hispanic	109	16.8%	1,118	25.5%
White (non-Hispanic	519	80.0%	3,117	71.1%

Decision Rules of Screening Tools

Students were identified as needing reading interventions based on decision rules, or cut points. Table 3 summarizes the decision rules for each screening tool. Using the CSAP (CTB/McGraw Hill, 2007, 2008) results on the reading portion, students who performed at the Proficient or Advanced performance levels were judged to not be in need of an intervention. Students who received a performance level of Partially Proficient were identified for Tier 2 interventions, and those students who received an Unsatisfactory performance level were identified as needing Tier 3 intervention. For the TOSCRF (Hammill et al., 2006) and the TOSWRF (Mather et al., 2004), students who performed at or below the 16th percentile were identified for Tier 3 intervention, while

those students performing above the 16th percentile but below the 40th percentile were identified for Tier 2 intervention.

Table 3

Decision Rules for Identification of Interventions

	Screening Tool		
	CSAP	TOSCRF	TOSWRF
No intervention	Advanced or Proficient performance level	At or above 40th percentile	At or above 40th percentile
Tier 2 intervention	Partially Proficient performance level	Below 40th percentile and above 16th percentile	Below 40th percentile and above 16th percentile
Tier 3 intervention	Unsatisfactory performance level	At or below 16th percentile	At or below 16th percentile

Descriptive Statistics of Screening Tools

Table 4 represents the frequency of categorization by intervention level for each of the three screening tools. The frequency rates using the TOSCRF and TOSWRF appear to be similar. However, 73.4% of the sample was identified as not needing an intervention using the CSAP as compared to 52.5% and 55.3% using the TOSCRF and

TOSWRF, respectively. The mean percentiles of the TOSCRF and TOSWRF are 44.45 and 48.55, respectively, and the standard deviations are 24.97 and 27.46, respectively.

Table 4

Description of Study Sample – Screening Tools

	n	Frequencies
CSAP levels		
Unsatisfactory	38	5.9%
Partially Proficient	135	20.8%
Proficient	408	62.9%
Advanced	68	10.5%
TOSCRF percentiles		
≥ 16th percentile	101	15.6%
< 40th percentile, >16th percentile	207	31.9%
≥ 40th percentile	341	52.5%
TOSWRF percentiles		
≥ 16th percentile	96	14.8%
< 40th percentile, >16th percentile	194	29.9%
≥ 40th percentile	359	55.3%

Analysis

Research Questions 1 and 2

The Wilcoxon Signed Ranks Test compared the classification of students based on the CSAP (CTB/McGraw Hill, 2007, 2008) as a universal screener to classification of students based TOSCRF (Hammill et al., 2006) and TOSWRF (Mather et al., 2004). Table 5 presents the results of the Wilcoxon test and indicates how students were categorized using CSAP and TOSCRF and CSAP and TOSWRF. The results indicate that good agreement did not exist between TOSCRF and TOSWRF and the CSAP (Z = -9.86, $p < .001$ and Z = -8.89, $p < .001$, respectively).

Table 5

Results of Wilcoxon Signed Ranks Test

	n	Frequencies
CSAP and TOSCRF		
Negative ranks	57	8.78
Positive ranks	217	33.43
Ties	375	57.78
		Z = -9.86, p = <.001
CSAP and TOSWRF		
Negative ranks	63	9.70
Positive ranks	201	30.97
Ties	385	59.32
		Z = -8.89, p = <.001
TOSCRF and TOSWRF		
Negative ranks	124	19.10
Positive ranks	109	16.79
Ties	416	64.09
		Z = -1.368, Asymp. Sig. (2 tailed) = .171

Research Question 3

To determine if the results of the three screening tools differ for students by gender, a Friedman Two-Way Analysis of Variance for related samples and a Chi-Square Test of Independence were completed. Table 6 presents the percent distribution of students for gender by test. A comparison of males versus females for the CSAP did not find any difference, $\chi^2(2) = 4.22$, $p = .12$. Comparison of males and females on the TOSWRF also was not significant, $\chi^2(2) = 2.01$, $p = .37$. However, a comparison of males and females on the TOSCRF was significant $\chi^2(2) = 8.22$, $p = .02$). Examination of the percent of students identified as at risk by gender shows that males were more often categorized as needing interventions than females, 51.4 % and 42.5%, respectively.

The results of the Friedman analyses comparing distributions across the screening tools for males and females separately found significant differences between CSAP and TOSCRF, $\chi^2(2)= 69.85$,$p < .001$ and between CSAP and TOSWRF, $\chi^2(2) = 50.13$, $p < .001$). Examination of the percent distributions indicates that the CSAP identifies fewer students of both genders as needing intervention at any level.

Table 6

Categorization of Interventions by Gender

	Percent of students categorized		
	CSAP	TOSCRF	TOSWRF
Male			
No intervention	70.6	48.6	53.3
Tier 2 intervention	22.2	32.5	30.3
Tier 3 intervention	7.2	18.9	16.4
Female			
No intervention	76.8	57.4	57.8
Tier 2 intervention	19.0	31.1	29.4
Tier 3 intervention	4.2	11.4	12.8

Research Question 4

To determine if the screening results of the three tools differed for students by Hispanic and White, non-Hispanic, a Friedman and Chi-square tests were conducted. Results are presented in Table 7. A comparison of Hispanic students versus White, non-Hispanic students using CSAP, TOSCRF or TOSWRF showed significant difference, $\chi^2(2) = 66.11$, $p < .001$; $\chi^2(2) = 21.028$, $p < .001$; and $\chi^2(2) = 15.945$, $p < .001$, respectively. Hispanic students appear to be more likely to be categorized as needing reading interventions than White, non-Hispanic students; however, this result could be due to differences in the sample sizes.

Table 7

Categorization of Interventions of Hispanic and White, non-Hispanic

	Percent of students categorized		
	CSAP	TOSCRF	TOSWRF
	Hispanic		
No intervention	43.1	39.4	43.1
Tier 2 intervention	41.3	31.2	30.3
Tier 3 intervention	15.6	29.4	26.6
	White, non-Hispanic		
No intervention	80.2	55.7	58.8
Tier 2 intervention	16.0	31.8	28.7
Tier 3 intervention	3.9	12.5	12.5

To determine if differences exist for categorization of Hispanic students or White, non-Hispanic students across the three screening tools, the Friedman test was completed. The distribution of Hispanic students across the three screening tools is not significantly different, ($\chi^2(2) = 3.71$, $p = .16$); however, distribution of White, non-Hispanic is significant, $\chi^2(2) = 122.07$, $p < .001$). This finding indicates that the distribution across categories for interventions for Hispanic students differs from that of White, non-Hispanic students. Again, this finding may be due to the larger sample size for White students, 80% of the sample, as compared to the sample size for Hispanic students, 17%.

Research Question 5

Testing for the effect of a gender by ethnicity interaction was conducted using Hierarchical Log Linear analyses. The interaction was found not to be significant for CSAP, TOSCRF and TOSWRF ($\chi^2_2 = 0.05$, $p = .98$; $\chi^2_2 = 3.54$, $p = .17$; $\chi^2_2 = 0.39$, $p = .82$, respectively).

Research Question 6

Finally, to determine if value is added when combining the TOSCRF and/or TOSWRF with the CSAP to categorize students for interventions, cross tabulation analysis was completed. The three screening tools identified the same students as not needing intervention with 51% agreement; students categorized as Tier 2 intervention with 18.5% agreement, and students categorized as needing Tier 3 intervention with 42% agreement.

Comparing CSAP and TOSCRF, the CSAP categorized 8.8% of students as needing an intervention when the TOSCRF indicated no intervention was needed. Conversely, the CSAP categorized 33.4% of students as not needing an intervention when the TOSCRF indicated an intervention was needed.

Comparing the CSAP with the TOSWRF, the CSAP categorized 9.7% of students as needing interventions when the TOSWRF indicated no interventions were needed. Conversely, the CSAP categorized 46.4% of students as not needing interventions when the TOSWRF indicated interventions were needed. This pattern implies that the hit rate for categorization using the CSAP is lower, resulting in higher false negatives as compared to either the TOSCRF or the TOSWRF. Therefore, value is added when the CSAP results are used in conjunction with either the TOSCRF or TOSWRF.

Finally, an additional analysis compared the categorization results of the TOSCRF with the TOSWRF. No significant difference was found between the TOSCRF and the TOSWRF ($p = .171$), meaning the results matched more frequently than did either of these screening tools with the CSAP.

Additional Analysis

Analysis, using and Chi-Square Test of Independence, was also completed to determine if differences existed by grade within each screening tool. Table 8 displays the results. No significance was seen within the CSAP by grade, $\chi^2(12) = 19.0$, $p = .09$; nor was there significance seen within the TOSWRF, $\chi^2(12) = 24.6$, $p = .033$. Significance was seen within the TOSCRF by grade, $\chi^2(12) = 16.6$, $p = .001$.

Table 8

Categorization of Interventions by Grade Level

	Percent of students categorized		
	CSAP	TOSCRF	TOSWRF
	Grade 5		
No intervention	62.9	46.8	41.9
Tier 2 intervention	27.4	33.9	38.7
Tier 3 intervention	9.7	19.4	19.4

	Grade 6		
No intervention	76.6	42.3	55.5
Tier 2 intervention	19.0	35.8	35.8
Tier 3 intervention	4.4	21.9	8.8
	Grade 7		
No intervention	69.7	50.6	44.9
Tier 2 intervention	24.7	30.3	32.6
Tier 3 intervention	5.6	19.1	22.5
	Grade 8		
No intervention	71.0	56.1	53.3
Tier 2 intervention	23.4	32.7	29.9
Tier 3 intervention	5.6	11.2	16.8
	Grade 9		
No intervention	67.4	59.8	67.4
Tier 2 intervention	27.2	37.2	17.4
Tier 3 intervention	5.4	13.0	15.2
	Grade 10		
No intervention	80.7	54.5	59.1
Tier 2 intervention	15.9	31.8	28.4
Tier 3 intervention	3.4	13.6	12.5

	Grade 11		
No intervention	82.4	62.2	62.2
Tier 2 intervention	8.1	29.7	25.7
Tier 3 intervention	9.5	8.1	12.2

Finally, the Friedman Two-Way Analysis of Variance for related samples was used to determine differences among the three screening tools by grade and Table 9 displays the results. Significance was seen across the three screening tools within each grade, with the exception of grade 9.

Table 9

Friedman Test Results for Comparison Across Tests Within Grade

Grade	n	χ^2	df	*p*
5	62	15.13	2	.001
6	137	52.77	2	.000
7	89	19.30	2	.000
8	107	16.50	2	.000
9	92	2.57	2	.277
10	88	23,66	2	.000
11	74	12.37	2	.002

Summary

Results indicate the CSAP as a screening tool to categorize students for tiered levels of intervention do not match the categorization results of the TOSCRF or TOSCWRF as screening tools. No inter-test differences in categorization results by gender or by ethnicity (Hispanic versus White, non-Hispanic) exist, however intra-test differences were found, particularly for the TOSCRF among females and Hispanic students, within each of the three screening tools. No significance was seen within each

screening tool by grade; there was, however significance across the three screening tools within each grade, with the exception of grade 9. Finally, results indicate that adding the TOSCRF or TOSWRF to a screening process that use the CSAP results can add value in that more students will be identified based on reading weaknesses.

Chapter Five

Discussion and Conclusion

This chapter presents a summary of the study and important conclusion implied by the data presented in Chapter Four. Implications for action and recommendations for further research are also discussed.

Summary of the Study

Despite the progress that has been made over the past 10 years in understanding and addressing the reading problems in elementary school students (McCardle & Chhabra, 2004), more than eight million adolescents continue to struggle with basic reading (Kamil, 2003). Studies indicate that these students can have both word level and comprehension level deficits. One of the keys to addressing the problem is early identification and intervention. Recently an instructional model that provides a system of tiered interventions to students at risk or who have developed academic or behavioral weaknesses, Response to Intervention (RTI), has been promoted nationally by some prominent researchers (Batsche et al., 2008; Chun & Witt, 2008; Fuchs & Fuchs, 2005; Johnson et al. 2006; Vaughn, 2005; Shinn, 2008; Sugai & Horner, 2007). RTI contains the key components of universal screening, progress monitoring, levels or tiers of intervention, and fidelity of implementation. These components together provide a system of high quality research-based instruction, monitoring of each student's performance as learning occurs, provision of preventions to those students at risk,

provision of interventions to those students not responding, and measurement of fidelity of implementation of the intervention.

The first step in providing such levels of reading intervention is to identify students who are at risk or who have developed weaknesses in reading despite receiving effective, scientific, research-based instruction. Universal screening is the process of assessing all students to distinguish those who are in need of assistance from those who are not in need, and then determining the level of intervention they need, Tier 2 or Tier 3. Tier 2 interventions are provided to students in addition to the core instruction and focus on the identified reading weaknesses. Tier 3 interventions are also additional to the core instruction and more intense in frequency and duration in order to address those students with the most significant weaknesses and those students who did not respond or responded poorly to Tier 2 instruction. Screening tools should be efficient, cost effective and be repeatable (Johnson et al., 2006). In addition, the screening tools should be accurate and have sensitivity and specificity rates of 80% to ensure that students who need interventions are not overlooked (Jenkins, 2003). Screening tools for use with secondary level students do not exist in abundance.

The purpose of this study was to examine three screening tools used at the secondary school level for identification of the need for tiered interventions. The first tool is a state assessment, the results of which are used to appropriately match the student with the needed level of intervention. The CSAP (CTB/McGraw Hill, 2007, 2008) results are readily available to schools, therefore making their use efficient and cost effective. The second screening tool, the TOSCRF (Hammill et al., 2006), measures the silent general reading ability of students ranging in age from 7 years 0 months to 18 years 11 months.

The TOSCRF is easily administered, has multiple forms for repeated testing and is low-cost. The third screening tool, TOSWRF (Mather et al., 2004), also assesses general reading ability by measuring word identification, fluency, and word comprehension of students ages 6 years 6 months through 17 years 11 months. The instrument also has multiple forms for repeated testing and is low-cost.

The key research questions addressed in this study were

1. How does using the results of the Colorado State Assessment Program as a universal screener to identify students in need of reading interventions compare with using the Test of Silent Contextual Reading Fluency?
2. How does using the results of the Colorado State Assessment Program as a universal screener to identify students in need of reading interventions compare with using the Test of Silent Word Reading Fluency?
3. Are the results of the three screening tools different for males and females?
4. Are the results of the three screening tools different for students of different race/ethnicity: Hispanic/White?
5. Do the three screening tools together add value to the process of identifying students in need of reading interventions?

As indicated by the research questions, not only was a comparison of the three tools examined, but also the results by gender and ethnicity (Hispanic versus White, non-Hispanic) were analyzed to determine if the results were significantly different for a these group depending on the screening tool used.

The study compared the results of a single sample of students with three assessments. Schools districts were asked to share data that had already been collected as

part of their participation in the CDOE initiated Rural Secondary Literacy Project. A total of 1,763 student files were collected and after excluding files with missing test results, 649 students across fifth through eleventh grade from three rural districts with total student enrollment under 2,000 comprised the sample dataset. Examination of ethnicity was limited to differences between Hispanic and White, non-Hispanic only, due to very small number of number of students who were American Indian or Alaskan Native, Asian, or Black non-Hispanic. Students were fairly evenly distributed across the grade level and gender; however, an over representation of White, non-Hispanic students and an under representation of Hispanic students did exist in the sample when compared to the combined ethnic distributions of the three participating districts. The CSAP was administered to students in the spring of 2007 and the TOSCRF and TOSWRF were administered to the same students the following fall. Using the CSAP performance levels, students were categorized as not needing intervention if they received an Advanced or Proficient performance level, needing Tier 2 intervention if they received a Partially Proficient performance level and needing Tier 3 if they received an Unsatisfactory performance level. Using the TOSCRF or TOSWRF, students were categorized as not needing intervention if the results were at or above the 40th percentile, needing Tier 2 if the performance was below the 40th percentile but above the 16th percentile, and needing Tier 3 if the results were at or below the 16th percentile.

Descriptive statistics were complied from the sample set, including frequencies for each of the three screening tests. Comparative analysis of the results of the three screening tools was completed using the Wilcoxon Signed Ranks Test. The Friedman Two-Way Analysis of Variance for related samples and Chi-Square Test of Independence

were completed by gender and ethnicity. Hierarchical Log Linear tests were run to determine if there was a gender by ethnicity interaction effect for each of the three screening tools.

Findings

As indicated in the review of the literature, many adolescents continue to struggle with reading as they progress through secondary schools. These adolescents can have both word level and comprehension weaknesses. A system of RTI calls for universal screening to determine which of those students are in need of tiered levels of reading interventions. In universal screening, a screening tool should be efficient, cost effective, and yield reasonably accurate results.

The results of this study indicate that a significant difference exists between the CSAP and each of the two nationally-normed reading screening tools. The TOSCRF and TOSWRF identified more than 20% additional students as needing reading interventions than the CSAP. It may be that solely using the CSAP as a screening tool could result in under identification of students who may be needing interventions based upon a nationally normed reading screening. A more accurate approach may be to combine the state assessment results with the TOSCRF and TOSWRF to make screening decisions. In addition, no significant difference was seen in results between the TOSCRF and TOSWRF; therefore, using only one of these screening tools would be more cost effective because there appears to be no benefit in using both.

This has significant implications for the field. It has been common practice to use the results of CSAP to target specific students for interventions in order to move more students from the Partially Proficient to the Proficient level in order to improve the

adequate yearly progress score of the school. Schools report the number of students proficient or advanced as a means of demonstrating the quality of the school. However this may be misleading, giving a false sense of success resulting in a number of students falling through the cracks. Unaddressed, these students continue to struggle as they are faced with increased complexity in content area classes, only to fail and eventually dropout. Thus, the district or school leadership must struggle with whether to use the CSAP results and provide less students with reading interventions which could save valuable resources or use a screening tool like the TOSCRF or TOSWRF and identify and provide a larger number of students with needing interventions. The latter would result in a larger number of students having the necessary skills to be successful in and beyond the public school system and financially contribute more to the economic state of our nation. This struggle becomes even more poignant for adolescents whose skills stay stagnant or decrease, while the time to address those weaknesses lessens.

Second, this study implies that although a state assessment for reading may reflect student performance according to the standards, these standards may not reflect the rigor demonstrated by reading research and such nationally normed reading assessments as the TOSCRF and the TOSWRF. Based on the state reading assessment, a number of students who were designated as proficient; however, on the nationally-normed assessments many of these same students demonstrated weak reading skills. Practitioners, teachers in particular, and parents should not assume that a student proficient on a state assessment is a proficient reader. In addition, if a student is identified as not needing interventions based on the results of state assessment, the decision could have adverse impact as the student moves from a state with less rigorous standards that align poorly to a national

assessment such as the National Assessment Education Program (NAEP)—the national measure of academic achievement—to a state that is more rigorous and closely aligned with the standard. A general agreement exists that a large number of state testing results do not align with the NAEP. In 2005, 86 % of eighth grade students performed at the proficient or advanced levels on the CSAP, however only 37% of eighth grade students performed at or above the proficient level in the NAEP (National Center for Education Statistics, 2005). If screening tools based upon local norms or a state's standards and benchmarks are utilized to determine which students receive academic assistance, then those students in states with less rigorous standards will continue to receive inadequate instruction rather than rigorous effective instruction matched to their needs, which is the foundation of RTI. Indeed, the issue of state standards and lack of rigor is at the forefront of today's education issues. In a recent issue of *Time* magazine (2009), Walter Isaacson, president and CEO of the Aspen Institute (an institute that provides a neutral and balanced venue for examination and action planning on critical issues such as education), summarized the current state of the K12 education system calling it "and incoherent jumble of state and local curriculum standards, assessments tools, tests, texts and teaching materials" (p. 1). Rather than rising to a high level of excellence, states have defined their standards lower in order to superficially satisfy federal demands of proficiency. The National Governor's Association, the Council of Chief State School Officers and Achieve, Inc. are studying on the concept of common national standards with internationally recognized rigor. This study occurs at a time when many are examining how we measure student performance across the nation.

No significant differences were observed by gender across the three tests. Interestingly but not surprising, more males were identified as needing more intensive intervention (Tier 3) than females. The National Assessment of Education Progress (NAEP) in 2007 showed females outscored their male counterparts by 10 points (NAEP, 2009). Another interesting finding was that females performed significantly higher compared to males on the TOSCRF. Females with adequate reading skills may have better contextual reading skills at the secondary level. In other words, they may have moved beyond word level meaning to more successful use of the context level of reading. However, more understanding of gender differences in these two tools include needs to be accomplished through further research to ensure the most accurate unbiased screening tools are available for use.

When examining Hispanic/White ethnicity, a significant difference was seen between the performance of Hispanic students and White, non-Hispanic students. More Hispanic students were identified as needing interventions than White students. This finding is consistent with the results of the NAEP, which demonstrated significantly lower performances by Hispanic students when compared to White, non-Hispanic students (NAEP, 2009). The finding is also consistent with the theories of discourse processing and research on literacy development of ELLs that suggest that a potential obstacle to adolescent reading comprehension is lack of language knowledge and experience (Kintsch, 1998; Kintsch, 2005), which is often the case with ELLs (Lesaux, & Geva, 2006; Geva, 2006). However, worth noting from this study, is the possibility of low performance by Hispanic students being a result of the sample size.

When examining performance related to grade level, the CSAP identified fewer students as needing reading interventions throughout the grade levels. CSAP under identifies students for reading interventions regardless of grade level. There did appear to be significance for the TOSCRF and TOSWRF across grade levels. In each case there was an increase in students identified as needing interventions between grades 5 and 6. This could be due to the developmental and instructional shift that occurs between elementary school and middle schools. Each of the three participating districts comprised elementary school as ending with fifth grade and middle school beginning with sixth grade. An instructional shift occurs, moving from learning to read to reading to learn content material. In addition, there is a developmental shift that occurs as students move to more inferential and analytical comprehension of contextual concepts. However, more research needs to be conducted to better understand the difference noted in this study.

Another noteworthy observation is in the categorization across screening tools with respect to Tier 2 identification. Less agreement occurred among tests for Tier 2 determination than among tests for Tier 3 or no intervention. This pattern seems to be consistent with what others have seen across the nation as schools implement RTI. As a technical assistance provider for a national RTI center, this researcher has received many more questions regarding who should be in Tier 2 and what interventions should be used for students in Tier 2 than Tier 3. Speculatively, practitioners may be struggling with the decision rules or cut points, the specific tools, and/or the implementation of a system of identification, which calls for the clear discrimination among academically weak students.

Limitations

Limitations to this study are present. First, there were a significant number of cases missing data from the original 1,763 submitted by the participating districts. Information on special education and English language learner status was sporadic. It was not clear as to the cause, however this brings attention to the nature of data reporting systems in smaller, rural districts. Analysis had to be conduced using Hispanic versus White, non-Hispanic rather than all ethnic subgroups as defined by the ethnicity coding required by the Colorado Department of Education.

In addition, although a large number of student results were collected, over half of the cases had missing data, which may lead to questions the fidelity of the implementation of the screening process. The missing data did not allow for examination of the performance differences on each screening assessment of the students identified as special education or at different levels of English proficiency, which are important to the nature of selecting appropriate unbiased screening tools. In addition, comparing the differences in performance on the three screening tools of students identified as limited English proficient with those who are designated as Fluent English proficient would assist in understanding and designing appropriate interventions for students whose English is a second language.

Second, the sample is composed of only three rural districts with relatively low minority representation as compared to the state of Colorado. Including other ethnic subgroups, Asian, American Native and Black, in a comparison would have given

additional value to the study. Given these limitations, the results of this study should be viewed cautiously and not generalized to a larger population.

Implications for Future Research

A number of implications for future research and practice emerge from this analysis. Districts and schools claim to be successfully implementing RTI to address the needs of students who exhibit reading difficulties, however if effectiveness is the goal of implementation more needs to be learned to guide implementation. The present study emphasizes some of the concepts decision makers need wrestle with when implementing an RTI system. Accurately identifying the right students for the right level of intervention is a multi-faceted task and this study examined only one small aspect of that task. Educational leaders need to be cognizant of the enormous responsibility placed on universal screening tools and decision criteria, and that using the data effectively to intervene and improve results for all students may be vital to student success.

This study examined the initial part of the screening process. In order to determine the true specificity and sensitivity rates of a screening tool, research needs to be extended beyond the initial identification. Research needs to be conducted that examines the students' responses to intervention as part of a validity process for the screening tools. The possibility exists that during the intervention, progress monitoring could demonstrate that some students who were initially identified making adequate progressing, and therefore may have been falsely identified. Such additional research could more accurately validate the level of specificity and sensitivity. In addition, conducting a second and third screening during the academic year could also help to identify any students who may have been missed in the initial identification as needing

interventions. A study with a larger sample size with representation in all ethnicities and that includes representation from urban and suburban districts should also be conducted.

More research needs to be conducted to examine the performance of specific subgroups, especially English language learners, on screening tools to determine the most unbiased accurate screening tool for reading. In addition, the intervention needs of these students are different and must be identified in order to accurately match the interventions to the students' needs. This is critical if a multi-system of prevention and intervention like RTI is to be successful in closing the reading achievement gap among students with diverse circumstances.

Additional research is needed to examine performance on screening tools at various grade levels. The grade level analysis completed in this study indicated some differences in identification across tests by grade, however conducting a longitudinal study with a cohort of students across grades is needed to further explore and understand these differences.

Finally, a screening tool needs to be efficient and cost effective. The TOSWRF and TOSCRF used in this study can be considered efficient in that each can be group administered in three minutes by a classroom teacher and the scoring is relatively uncomplicated. Informal conversations from a coach assigned to two of the schools participating in the rural literacy project indicated that administration and scoring for one student took no more than ten minutes. The cost of the TOSCRF materials for 1750 students, approximately the number of files collected from the three districts participating in this study, was approximately $2.15 per student. The cost of the TOSWRF materials for the same number of students was $0.88 per student. This does not include the human

costs to order, train, administer, collect and share the results; however, these expenses are far less than utilizing individual assessments or large assessment systems. For example, AIMSweb® can be used for screening through eighth grade for a cost ranging form $2.00 to $4.00 per student (Edinformation, Inc., 2009) and Measures of Academic Progress assesses students in grades 3 through 10 for approximately $12.50 per student. There are additional costs for technology related to both these assessments. Research to examine the cost effectiveness of screening tools is needed.

Bibliography

Alliance for Excellent Education (2006). *Fact sheet: Adolescent literacy.* Washington, DC: Author.

Anatasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). New York: Macmillan.

Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning, decision making, and the identification of children in need of services. *Journal of Psychoeducational Assessment*, *23*(4), 362–380.

Ardoin, S. P., Witt, J. C., Suldo, J. C., Connell, J. E., Koenig, J. L., Resetar, J. L., et al. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review*, *33*(2), 218–233.

Bacon, S. (2005). Reading coaches: Adapting an intervention model for upper elementary and middle school readers. *Journal of Adolescent & Adult Literacy*, *48*(5), 416–427.

Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, *42*(4), 223–235.

Barnes, K. E. (1982). *Preschool screening: The measurement and prediction of children at-risk*. Springfield, IL: Charles C. Thomas.

Batsche, G., Castillo, J., Dixon, D., & Forde, S. (2008). Best practices in linking assessment to intervention. In Thomas, A. & Grimes, J. (Eds.) *Best practices in school psychology V* (pp. 177-93). Bethesda, MD: National Association of School Psychologists.

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

Bell, S. M., McCallum, S., Burton, B., Gray, R., Windingstad, S. & Moore, J. (2007). Concurrent validity of the test of silent word reading fluency. *Assessment for Effective Intervention, 31*(3), 1–9.

Bell, S. M., McCallum, S., Kirk, E. R., & McCane-Bowling, S. (2007). Investigation of the psychometric attributes of the test of silent contextual reading fluency. *Assessment for Effective Intervention 33*(1), 39–46.

Bergan, J. R. (1977). *Behavioral consultation*. Columbus, OH: Charles Merrill.

Bergan, J. R., & Kratochwill, T. R. (1990). *Behavioral consultation and therapy.* New York: Plenum Press.

Biancarosa, C., & Snow, C. E. (2006). *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.

Bollman, K. A., Silberglitt, B., & Gibbons, K. A. (2007). The St. Croix River education district model: Incorporating systems-level organization and a multi-tiered problem-solving process for intervention delivery. In M. Burns, S. Jimerson, & A.VanDerHeyden (Eds). *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 319-330). New York: Springer Science + Business Media, LLC.

Buly, M. R., & Valencia, S. (2003). *Meeting the needs of failing readers: Cautions and considerations for state policy.* Seattle: University of Washington, Center for the Study of Teaching and Policy.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment*, *23*(4), 381–394.

Callender, W. A. (2007). The Idaho results-based model: Implementing response to intervention statewide. In M. Burns, S. Jimerson, & A.VanDerHeyden (Eds). *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 331-342). New York: Springer Science + Business Media, LLC.

Calhoun, M. B., & Fuchs, L. S. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial and Special Education, 24*(4), 235–45.

Catts, H. (2006). *School-wide screening.* Presentation at the National SEA Conference on Response to Intervention: Integrating RTI within the SLD Determination Process, Kansas City, MO. Retrieved December 11, 2008, from http://www.nrcld.org/sea/presentations_worksheets/screening/Catts_screening.pdf

Catts, H. W., Hogan, T. P., & Adolf, S. M. (2005). Developmental changes in reading and reading disabilities. In H. W. Catts, & A. G. Kamhi (Eds.). *The connections between language and reading disabilities* (pp. 25–40). Mahwah, NJ: Lawrence Erlbaum Associates.

Chun, M., & Witt, J. (2008). *System to enhance educational performance*. Retrieved July 7, 2008, from http://www.isteep.com/index.html

The College Board (2005). *ACCUPLACER- OnLine The proven innovator for enhancing retention and student success.* New York, NY: Author.

Collins, J. L., Small, M. L., Kane., L., Pateman., B. C., Gold., R. S., & Kolbe, L. J. (1995). School Health Education. *Journal of School Health, 65*(8), 302–311.

Colorado Department of Education [CDOE] (2009). *The Colorado State Assessment Program.* Denver, CO: Author. Retrieved March 9, 2009 from http://www.cde.state.co.us/cdeassess/documents/csap/usa_index.html

Conderman, G. & Petersen, T. (2007). Avoid the tutor trap. *Intervention in School & Clinic 42*(4), 234–238.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity on psychological tests. *American Psychologist, 52*(4), 281–302.

CTB/McGraw-Hill for the Colorado Department of Education (2007). *The Colorado state assessment program technical report.* Denver, CO: Author. Retrieved March 9, 2009 from http://www.cde.state.co.us/cdeassess/documents/reports/2007/CSAPTechnicalReport2007Text_Final.pdf

CTB/McGraw-Hill for the Colorado Department of Education (2008). *The Colorado state assessment program technical report.* Denver, CO: Author. Retrieved March 9, 2009 from http://www.cde.state.co.us/cdeassess/documents/reports/2008/CSAPTechnicalReport2007Text_Final.pdf

Deno, S. L., & Mirkin, P. H. (1977). *Data-based program modification: A manual*. Minneapolis, MN: The Council for Exceptional Children.

Denton, C. A., & Wexler, J. (2006, April). *Older students with reading difficulties: The search for solutions*. Paper presented at the annual meeting of the Council for Exceptional Children, Salt Lake City, UT.

Deshler, D., & Hock, M. (2007). Adolescent literacy: Where we are, where we need to go. In A. K. Billman, K. Perry, M. Pressley, K. Reffitt, J. M. Reynolds (Eds.). *Shaping literacy achievement: research we have, research we need* (pp. 98-128). New York: Guilford Press.

Duffy, H. (2007). *Meeting the needs of significantly struggling learners in high school: A look at approaches to tiered intervention*. Washington, DC: U.S. Department of Education, National High School Center.

Durlak, J. A. (1997). *Successful prevention programs for children and adolescents*. New York: Plenum Press.

Edformation, Inc. (2009). AIMSweb® pricing. Retrieved May 9, 2009, from http://www.aimsweb.com/upload/pdfs/aimsweb_pricing_813-4.pdf

Espin, C., & Halverson, N. (1999). Criterion-related validity of curriculum-based measures in writing for secondary students. *Reading and Writing Quarterly, 15*(1), 5-27.

Epsin, C., Scierka, B. J., & Skare, S. (1999). Criterion-related validity of curriculum-based measures for writing for secondary students. *Reading and Writing Quarterly, 15*(1), 5–27.

Fisher, D. (2001). "We're moving on up": Creating a schoolwide literacy effort in an urban high school. *Journal of Adolescent & Adult Literacy, 45*(2), 92–101.

Foorman, B. R., Fletcher, J. M., Frances, D. J. Carlson, C. D., Chen, D., Mouzaki, A., Schatschneider, C., Wristers, K., & Taylor, R. (1998). Technical Report Texas Primary Reading Inventory Technical (1998 Edition). Houston, TX: Center for Academic and Reading Skills, University of Houston.

Foorman, B. R., Winikates, D., Mehta, P., Schatschneider, C., & Fletcher, J. M. (1997). Early interventions for children with reading disabilities. *Scientific Studies of Reading, 1*(3), 255–276.

Frankenberg, W. (1974). Selection of diseases and tests in pediatric screening. *Pediatrics, 54*(5), 612–616.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J., & Hamlett, C. L. (2005). Responsiveness to intervention: Preventing and identifying mathematics disability. *Teaching Exceptional Children, 37*(4), 60–63.

Fuchs, D., & Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice, 22*(2), 129–136.

Fuchs, D., & Fuchs, L. S. (2005). Responsiveness-to-intervention: A blueprint for practitioners, policymakers, and parents. *Teaching Exceptional Children, 38*(1), 57–61.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disabilities by responsiveness to instruction: Specifying measures and criteria. *Learning Disability Quarterly, 27*(4), 216–228.

Fuchs, L. S., & Fuchs, D. (2006). Implementing responsiveness-to-intervention to identify learning disabilities. *Perspectives on Dyslexia, 32*(1), 39–43.

Geva, E. (2006). Second language oral proficiency and second language literacy. In D. August, & T. Shanahan (Eds.), *Developing literacy in second-language learners* (pp 123-139). Mahwah, NJ: Lawrence Erlbaum Associates.

Girshick, M. A. (1954). An elementary survey of statistical decision theory. *Review of Educational Research, 24*(5), 448–466.

Good, R. H., III, Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*(3), 257–288.

Graves, A., Block, N., & Brandon, R. (2009, April). *Response to intervention in sixth grade: Two studies of Tier II instruction*. Presented at the Council for Exceptional Children Annual Conference, Seattle, WA.

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15*(4)*,* 198–205.

Guion, R. M. (1974). Open a new window: Validities and values in psychological measurement. *American Psychologist, 29*(5)*,* 287–296.

Gulliksen, H. (1950). Intrinsic validity. *American Psychologist, 5*(10), 511–517.

Gunn, B., Biglan, A., Smolkowski, K., & Ari, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *Journal of Special Education, 34*(2), 90–103.

Gunn, B., Smolkowski, K., Biglan, A., & Black, C. (2002). Supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *Journal of Special Education, 36*(2), 69–79.

Hammill, D. D., Weiderholt, J. L., & Allen, E. A. (2006). *The test of silent contextual reading fluency examiner's manual.* Austin, TX: Pro-ed, Inc.

Harber, J. (1981). Evaluation utility in diagnostic decision making. *Journal of Special Education, 15*(4), 413–428.

Hock, M. F., Brasseur, I. F., Deshler, D. D., Catts, H. W., Marquis, J. G., Mark, C. A., & Stribling, J. W. (2009). What is the reading component skill profile of adolescent struggling readers in urban schools? *Learning Disabilities Quarterly, 32*(1), 21-38.

Hughes, C., & Douglas, D. D. (2009). *Universal screening within a response-to-intervention model*. Retrieved on January 2, 2009, from http://www.rtinetwork.org/Learn/Research/ar/Universal-Screening-Within-a-Response-to-Intervention-Model

Isaacson, W. (2009). How to raise the standard in America's schools. *Time,* April 27, 2009. Retrieved on April 25, 2009 from http://www.time.com/time/nation/article/0,8599,1891468,00.html

Jenkins, J. G. (1946). Validity for what? *Journal of Consulting Psychology*, *10*(2), 93–98.

Jenkins, J. R. (2003). *Candidate measures for screening at-risk students*. Paper presented at the NRCLD Response to Intervention Symposium, Kansas City, MO. Retrieved January 4, 2009, from http://www.nrcld.org/symposium2003/jenkins/index.htm

Jenkins, J. R., Peyton, J. A., Sanders, E. A., & Vadasy, P. F. (2004). Effects of reading decodable texts in supplemental first-grade tutoring. *Scientific Studies of Reading, 8*(1), 53–85.

Johnson, B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative, and mixed approaches* (2nd ed.). Boston: Pearson Education, Inc.

Johnson, E., Mellard, D. F., & McKnight, M. A. (2006). *Responsiveness to intervention (RTI): How to do it*. Lawrence, KS: National Research Center on Learning Disabilities.

Johnson, E. S., & Smith, L. (2008). Implementation of response to intervention at middle school. *Reading Improvement, 41*(1), 46-52.

Kamil, M. (2003). *Adolescents and literacy reading for the 21st century.* Washington, DC: Alliance for Excellent Education.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Elective classroom and intervention practices: A Practice Guide* (NCEE #2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved on April 4, 2009, from http://ies.ed.gov/ncee/wwc

Ketterlin-Geller, L., McCoy, J. D., Twyman, T., & Tindal, G. (2006). Using a concept maze to assess student understanding of secondary-level content. *Assessment for Effective Intervention, 31*(2), 39–50.

Kintsch, E. (1998). *Comprehension: A paradigm for cognition.* Cambridge, United Kingdom: Cambridge University Press.

Kintsch, E. (2005). Comprehension theory as a guide for the design of thoughtful questions. *Topics in Language Disorders, 25*(1) 51–64.

Kovaleski, J. F., Gickling, E. E., Morrow, H., & Swank, H. (1999). High versus low implementation of instructional support teams: A case for maintaining program fidelity. *Remedial and Special Education*, *20*(3), 170–183.

Leach, J. M., Scarborough, H. S., & Rescoria, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95*(2), 211-224.

Lee, J., Grigg, W., & Donahue, P. (2007). *The Nation's Report Card: Reading 2007* (NCES 2007–496). Washington, DC: U.S. Department of Education National Center for Education Statistics, Institute of Education Sciences.

Lesaux, N., & Geva, E. (2006). Synthesis: Development of literacy in language-minority students. In D. August, & T. Shanahan (Eds.), *Developing literacy in second-language learners* (pp. 53–74). Mahwah, NJ: Lawrence Erlbaum Associates.

Lichtenstein, R., & Ireton, H. (1984). *Preschool screening: Identifying young children with developmental and educational problems*. Orlando, FL: Grune & Stratton, Inc.

Mackinac Center for Public Policy. (2008). Into and beyond MEAP. Retrieved on May 9, 2009, from http://www.educationreport.orglpubs/mer/article.aspx?id=9993

Marston, D., Muyskens, P., Lau, M., & Canter, A. (2003). Problem-solving model for decision making with high-incidence disabilities: The Minneapolis experience. *Learning Disabilities Research & Practice*, *18*(3), 187–200.

Mather, N., Hammill, D. D., Allen, E. A., & Roberts, R. (2004). *Test of silent word reading fluency examiner's manual*. Austin, TX: PRO-ED, Inc.

McCardle, P., & Chhabra, V. (2004). *The voice of evidence in reading research.* New York: Brookes Publishing.

McMaster, K. N., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to non-responders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71*(4), 445–463.

Meisels, S. J. (1985). *Developmental screening in early childhood: A guide.* Washington, DC: National Association for the Education of Young Children.

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

Mercer, C. D., Campbell, K. U., Miller, M. D., Mercer, K. D., & Lane, H. B. (2000). Effects of reading fluency intervention for middle schoolers with specific learning disabilities. *Learning Disabilities Research & Practice, 15*(4), 179–189.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*(11), 1012–1027.

Morris, D., Tyner, B., & Perney, J. (2000). Early steps: Replicating the effect of a first-grade reading intervention program. *Journal of Educational Psychology, 92*(4), 681–693.

National Assessment of Education Progress (2009). The *nation's report card: Reading report 2007.* Retrieved on April 21, 2009 from http://nationsreportcard.gov/reading_2007/r0008.asp

National Center for Educational Statistics (2004). Language minorities and their educational and labor market indicators-Recent trends, NCES 2004–09. Washington, DC: US Department of Education, Author.

National Center for Educational Statistics (20050. *Mapping 2005 state proficiency standards onto the NAEP scales* NCES 2007-482. Washington, DC: US Department of Education, Author. US Government Printing Office.

National Center on Response to Intervention. (2008). *What is RTI?* Retrieved November 5, 2008 from http://www.rti4success.org

National Reading Panel. (2000). *Reports of the subgroups* (NIH Publication No. 00-4654). Washington, DC: National Institutes of Health.

O'Connor, R. E., Harty, K. R., & Fulmer, D. (2005). Tiers of intervention in kindergarten through third grade. *Journal of Learning Disabilities, 38*(6), 532–538.

Papalewis, R. (2004). Struggling middle school readers: Successful, accelerating intervention. *Reading Improvement, 41*(1), 24–37.

Perie, M., Grigg, W. S., & Donahue, P. L. (2005). *The nation's report card: Reading 2005* (NCES 2006-451). Washington, DC: U.S. Government Printing Office. Retrieved March 30, 2009 from http://nces.ed.gov/nationsreportcard/pdf/main2005/2006451.pdf

Peterson, D. W., Prasse, D. P., Shinn, M. R., & Swerdlik, M. E. (2007). The Illinois flexible service delivery model: A problem-solving model initiative. In M. K. Burns, S. R. Jimerson, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 300–318). New York: Springer.

President's Commission on Excellence in Special Education (2002). *A new era: Revitalizing special education for children and their families.* U.S. Department of Education contract No. ED-02-PO-0791. Washington, DC: Diskey & Associates, LLC.

Santa, C. M., & Hoien, T. (1999). An assessment of early steps: A program for early intervention of reading problems. *Reading Research Quarterly, 34*(1), 54–79.

Shinn, M., & Phillips, M. (2008, February). *Implementing response to intervention (RTI) in a problem-solving model at the secondary level*. Paper presented at the Annual Conference of the National Association of School Psychologists, New Orleans, LA.

Shinn, M. (2008). Implementation in secondary schools. In S. Fernley, S. D. LaRue, & J. Norlin (Eds.). *What do I do when: Answer book on RTI.* (pp. 8:1–17). Arlington, VA: LRP Publications.

Shinn, M., & Phillips, M. (2008, February). *Implementing response to intervention (RTI) in a problem-solving model at the secondary level*. Workshop presented at the National Conference of National Association of School Psychologists, New Orleans, LA.

Slavin, R. E., Karweit, N. L., & Madden, N. A. (1989). *Effective programs for students at risk*. Needham Heights, MA: Allyn & Bacon.

Sugai, G., & Horner, R. (2007). *Response to intervention logic and positive behavior support* [PowerPoint slides]. Retrieved on May 8, 2008 from http://www.pbis.org/main.htm

Torgeson, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency.* Lake Oswego, OR: Northwest Evaluation Association.

Twyman, T., & Tindal, G. (2007). Extending curriculum-based measurement into middle/secondary schools: The technical adequacy of the concept maze. *Journal of Applied School Psychology, 24*(1), 49–67.

U.S. Census Bureau. (2006). *Definition: Urban and rural.* Washington, DC: Author. Retrieved March 8, 2009 from https://ask.census.gov/

U.S. Department of Education (2004). *U.S. Department of Education's commentary and explanation about proposed regulations for IDEA 2004.* Washington, DC: Author.

Vaughn, S. (2005). *Interpretation of the 3-tier framework*. Retrieved May 9, 2008 from http://www.texasreading.org/utcrla/materials/3tier_letter.asp

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to intervention as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69*(4), 391–409.

VanDerHeyden, A., Witt, J., & Gilbertson, G. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45*(2), 225–256.

VanDerHeyden, A., & Witt. J. (2005). Quantifying context in assessment: Capturing the effect of base rates on teacher referral and a problem-solving model of identification. *School Psychology Review, 34*(2), 161–183.

Van Hook, J. (2008). *The reliability and validity of screening measures in reading*. Doctoral Dissertation, Louisiana State University.

Wald, A. (1950). *Statistical decision functions*. New York: John Wiley and Sons.

Wilson, J. M. G., & Jungner, G. (1968). *Principal and practices of screening for disease.* Public Health Papers No. 34, 1–163. Geneva, Switzerland: World Health Organization.

Witt, J., & VanDerHeyden, A. (2007). The system to enhance educational performance (STEEP): using science to improve achievement. In M. Burns, S. Jimerson, & A.VanDerHeyden (Eds.). *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 343–353). New York: Springer Science + Business Media, LLC.

Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests-Revised*. Circle Pines, MN: American Guidance Service.

Woodcock, R. W., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Riverside, CA: Riverside.

Zigmond, N. (1990). Rethinking secondary school programs for students with learning disabilities. *Focus on Exceptional Children 23*(1): 1–24.