

Developing a Teacher Rating Scale of Preschool Student Behavior for Use in an
RTI Decision-Making Framework

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Braden A. Schmitt

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Scott McConnell, Theodore Christ, Advisers

June 2014

Acknowledgements

I would like to thank and acknowledge the following colleagues, friends, and family for contributing to the ideas, skills, and other supports that lead to the completion of my dissertation:

my wonderful husband, Joseph Horton, for his patience and reminders to “notice“;

my parents, for cheering me on at every step;

my adviser and mentor, Scott McConnell, for his feedback, helpful diagrams, and stories about chickens and samurai;

Tracy Bradfield, Alisha Wackerle-Hollman, Michael Rodriguez, and the rest of the CRTIEC research team, for providing me with 5 years of fantastic scholarly teaching and experience;

Mandy Besner, Kate Clayton, Shawna Petersen-Brown, Megan Rodriguez, Cynthia Conner, and Laura Potter for keeping me sane and keeping me company during those 5 years;

the childcare centers, teachers, and children who participated so willingly in this project;

Karen Cadigan, for first convincing me that I was a school psychologist at heart.

Dedication

For John Yoakam, who first encouraged me to pursue a doctoral degree, who inspired me with his own academic accomplishments, and who supported me once I got there.
Thank you, John.

Abstract

The purpose of the present study was to evaluate a newly-developed measure of preschool learning-related behaviors, designed for the purpose of enhancing identification of the most appropriate tier of intervention for early literacy instruction, in part as a function of behavior characteristics of the student. Reliability and validity evidence were collected and evaluated. This included examination of internal consistency, test-retest reliability, and interrater reliability. Validity evaluation included the collection of criterion-related validity evidence, convergent and discriminant validity evidence, and an evaluation of the capacity of the measure to add value to a decision-making framework within a model of preschool Response to Intervention (RTI). The measure showed strong reliability and promising evidence regarding its capacity use teacher ratings of preschool learning-related behaviors to help identify the most appropriate tier of early literacy intervention. The limitations of the present study and implications for future research are discussed.

Table of Contents

Acknowledgements	i
Dedication	ii
Abstract	iii
Table of Contents	iv
List of Tables	v
List of Figures	vi
Introduction	1
Literature Review.....	6
Methods.....	36
Results	47
Discussion	58
References	72
Appendix A : CRTIEC Teacher Questionnaire (Sample A Version).....	85
Appendix B : CRTIEC Teacher Questionnaire Packet (Sample B Version)	86

List of Tables

Table 1. Demographic Information for Samples A and B.....38

Table 2. Descriptive Statistics by Measure.....47

Table 3. Mean and Standard Deviation for CTQ Total Scores.....47

Table 4. Descriptive Statistics for Week 1 and Week 2 CTQ Score.....48

Table 5. Descriptive Statistics for CTQ Total and BASC Total Scores.....51

Table 6. Correlations between Early Literacy Measures, Behavior Measures, and CTQ.....52

Table 7. Testing the Differences in Correlations for Like and Unlike Measures.....53

Table 8. Means, Standard Deviations for Groups by Age and Gender.....54

Table 9. Classification of Students into tiers by TOPEL, IGDIs and IGDIs+CTQ.....55

Table 10. Crosstabs Analysis for IGIDIs only and IGDI+CTQ.....58

List of Figures

Figure 1. CRTIEC Decision-making framework.....	23
Figure 2. Distribution of CTQ Total Scores for Samples A and B Combined.....	48
Figure 3 Correlation between CTQ scores at Week 1 and Week 2.....	50
Figure 4 Correlation between Rater A and Rater B.....	51
Figure 5 Correlation between CTQ Total Scores and BASC Total Scores.....	52
Figure 6 Scatterplot of CTQ total scores and IGDI Rhyming total scores.....	56
Figure 7 Scatterplot of CTQ total scores and IGDI First Sounds total scores.....	57

Chapter 1

Introduction

Early academic success, particularly in the area of early literacy, has been shown to be predictive of more positive educational trajectories later in schooling (National Early Literacy Panel, 2008; Scarborough, Dobrich, & Hager, 1991; Whitehurst & Lonigan, 1998). Students who do not develop certain literacy skills early on are at an increasing disadvantage for never fully developing adequate reading skills (Scarborough, 2001; Torgesen, Rashotte, & Alexander, 2001; Torgesen, 2002).

Researchers have identified four key areas of early literacy skills that are important to the development of later reading ability. These include oral language, phonological awareness, alphabet knowledge, and comprehension (Dunst, Trivette, Masiello, Roper, & Robyak, 2006; Senechal, LeFevre, Smith-Chant, & Colton, 2001; Whitehurst & Lonigan, 1998). In addition to these domains, some researchers have highlighted the importance of the role of preschool learning-related behaviors in the acquisition of academic skills (Bierman, Torres, Domitrovich, Welsh, & Gest, 2009; DuPaul, McGoey, Eckert, & VanBrakle, 2001; McClelland & Morrison, 2003; McWayne, Fantuzzo, & McDermott, 2004; Normandeau & Guay, 1998). Preschool learner behaviors have been shown to influence response to interventions as much as some early literacy skills (Nelson, Benner, & Gonzalez, 2003). A subset of preschool learning-related behaviors has shown a particularly important relationship to the

acquisition of early literacy skills; these behaviors are ones specific to participating in academic activities and are distinct from more general social skills (McClelland & Morrison, 2003). Despite the evidence regarding this relationship, the importance of preschool learner behavior is sometimes overlooked or addressed separately from research on the acquisition of early literacy skills.

Statement of the Problem

Preschool children come to the classroom with a wide range of abilities and behavioral needs. In order to maximize student response to instruction, students should receive instructional support that matches their individual level of need. One way to address this variation in student needs is Response to Intervention (RTI), a multi-tiered model that matches students to instructional conditions based on the results of systematic assessment (Jimerson, Burns, & VanDerHeyden, 2007). Given what is known about the importance of learning-related behaviors, one might assume that preschool student behavior needs should be included in the decision-making processes of RTI models. If learning-related behaviors are an important predictor of response to interventions, it follows that educators may need to assess students in this domain as well.

The inclusion of behavior and other learner characteristics in RTI decision-making models has been explored at the elementary level, and has aided decision-makers in making more accurate predictions about student response to intervention. (Compton, Fuchs, & Fuchs, 2006; Compton et al., 2010; Compton et al., 2012; Speece et al., 2011; VanDerHeyden, 2011). However, despite

recognition of the importance of learning-related behaviors to early learning, early research on preschool models of RTI has not yet incorporated these behaviors into instructional decision-making processes (Greenwood et al., 2011).

The nature of RTI requires that its component assessments possess certain characteristics in order to be feasible for use within the model. These characteristics include being psychometrically robust enough to make individual-level decisions about students who may require more intensive intervention, and being logistically feasible (i.e. brief and easy to administer) in order for early educators to easily use them with many students (Greenwood et al., 2008; Wackerle-Hollman, Schmitt, Bradfield, McConnell, & Rodriguez, 2013). While robust measures of preschool-student characteristics are available, many of the presently-available published measures of preschool student behavior tend to be too lengthy for feasibility within an RTI model. In addition, the available measures were not designed specifically for the purpose of identifying student characteristics that might influence response to instruction under different RTI conditions and may capture information that might not pertain directly to these types of decisions (Achenbach & Rescorla, 2000; Gresham & Elliot, 1990; Merrell, 2002; Reynolds & Kamphaus, 2010). If a measure of preschool student behavior is to be incorporated into an RTI decision-making model, the characteristics dictated by an RTI model likely need to be addressed and accommodated by that measure. Given this, there appears to be a need for an assessment of preschool learning-related behavior designed specifically for the

purpose of identifying students for different intervention conditions as part of an RTI decision-making framework.

Study Purpose

The present study is premised on the relatively simple argument that assessment of a particular set of learning-related behaviors has importance and usefulness within an RTI decision-making framework, in addition to content-related assessment of early literacy skills. The purpose of the present study was to evaluate a newly-developed measure of preschool learning-related behaviors, designed for the purpose of identifying the most appropriate tier of intervention for early literacy instruction as a function in part of behavior characteristics of the student. Reliability and validity evidence were examined in this study to make an overall evaluation of the support for including the assessment of learning-related behaviors as a part of an early childhood RTI decision-making framework.

1. What is the extent of the reliability evidence to support the CRTIEC Teacher Questionnaire (CTQ) as a consistent measure of preschool student learning-related behaviors?
 - a. What is the level of internal consistency (alpha) the CTQ?
 - b. How consistent are the ratings that teachers provide for the same child at two different times points?
 - c. How consistent are ratings provided by two different teachers for the same student?
2. What is the extent of the validity evidence to support the use of the CRTIEC Teacher Questionnaire (CTQ) as a measure of preschool learning-related behaviors to help inform decisions about tier assignment?

- a. What is the evidence that the CTQ sufficiently captures learning-related behaviors as demonstrated by a correlation between the CTQ and an already established measure of a similar domain?
- b. What is the extent of the convergent and discriminant evidence to indicate that the CTQ has the capacity to capture learning-related behaviors rather than constructs that it does not purport to measure, as measured by correlations between the CTQ and measures of Phonological Awareness?
- c. What is the evidence of the CTQ's capacity to add value to an early childhood RTI model by incorporating information about student learning-related behaviors in order to meaningfully influence tier assignment?

Chapter 2

Literature Review

Introduction

Students come to the classroom with widely varying skills and abilities; some exceed our initial expectations while others struggle to demonstrate basic skills. Providing instruction to students with such varying needs can present a challenge for school-age and preschool educators alike. Response to Intervention (RTI) is one model that has been proposed as a method for systematically assessing individual student needs and matching students to the appropriate level and intensity of instruction in order to enhance student outcomes (Jimerson et al., 2007). RTI is a multi-tiered model, with each successive tier providing more intensive and individualized instruction, along with more frequent progress monitoring (Fuchs & Fuchs, 2006). RTI advocates suggest that the benefits of an RTI model are its ability to provide educators with an efficient means for screening students, monitor their progress, and provide at-risk students with more intensive intervention before a persistent long-term pattern of failure emerges (Vaughn & Fuchs, 2003).

To be implemented successfully, an RTI model requires a measurement system that is quick, easy to implement, and psychometrically robust enough to make individual-level student decisions (Greenwood et al., 2008; Wackerle-Hollman et al., 2013). In addition to adequate measurement, an argument can be made that successful implementation of RTI requires a set of decision rules for how to optimally select students for candidacy in the different tiers of

intervention (Compton, Fuchs, Fuchs, & Bryant, 2006). While some RTI models select students for tiers based on assessment of one domain (i.e. reading fluency), some researchers have proposed that RTI decision-making frameworks include assessment of more than one domain, including skill level in different academic areas, rate of progress, and student behavior problems or other academic risk factors (Compton, Fuchs, & Fuchs, 2006; Compton et al., 2010; Compton et al., 2012; Speece et al., 2011; VanDerHeyden, 2011).

Collecting assessment data from different domains—such as academic skills and behavior—can help decision-makers make more accurate predictions about which students will or will not make adequate progress under typical instructional conditions (Compton et al., 2012). This then allows educators to change conditions and add supports for students who they predict may not make adequate progress under the current conditions (Compton et al., 2012; Speece et al., 2011). The evidence to support the use of such decision-making frameworks (by collecting evidence in multiple domains) has been gathered primarily in the school-age literature; support for this concept with preschool students is still limited (Compton, Fuchs, & Fuchs, 2006; Compton et al., 2010; Compton et al., 2012; Speece et al., 2011; VanDerHeyden, 2011). However, there is evidence that assessing preschool student learning-related behavior can be beneficial to these types of decisions (Nelson et al., 2003).

The following literature review will present an argument for the inclusion of assessments of certain behaviors in an RTI model for preschool. Specifically, it

will support the argument that teacher ratings of a particular type of student behaviors have a place in an RTI decision-making framework for preschool students. In addition, this review will summarize the necessary considerations for the development of a measure of student behavior for use in a preschool RTI decision-making framework.

Background

In the field of early education, one area, early literacy, has been identified as a keystone skill for preschool development. Within this domain, distinct trajectories have been identified for children who do and do not develop these skills prior to school entry or early in their education (National Early Literacy Panel, 2008; Scarborough et al., 1991; Whitehurst & Lonigan, 1998). Children who fail to adequately develop literacy skills are increasingly at a disadvantage as they progress through schooling. These children are more likely to miss out on opportunities to learn, never fully develop average level reading skills, and are more likely to develop reading disabilities (Scarborough, 2001; Torgesen et al., 2001; Torgesen, 2002).

Several important skill domains have been identified as critical to early literacy development (Dunst et al., 2006; Senechal et al., 2001; Whitehurst & Lonigan, 1998). Briefly, these domains are (a) oral language, or expressive and receptive vocabulary (Dunst et al., 2006); (b) phonological awareness, or the ability to detect and manipulate words at the level of the smallest units of spoken language (phonemes) (Anthony, Williams, McDonald, & Francis, 2007); (c)

alphabet knowledge and concepts about print, or the ability to recognize and produce letter names and sounds and the understanding of the conventions of written English (Strickland & Shanahan, 2004); and (d) comprehension, or the ability to gain information and draw inference from written and/or spoken language (Snow, Burns, & Griffin, 1999). Given their importance in literacy development, these domains are frequently a primary focus of instruction in early education.

One of these domains, phonological awareness (PA), has been shown to hold particular importance in later reading development (Anthony & Lonigan, 2004; Muter, Hulme, Snowling, & Stevenson, 2004). The process by which children learn PA skills differs somewhat from the other domains of early literacy (Lundberg, Frost, & Petersen, 1988). Whereas preschool students may acquire vocabulary or alphabet knowledge in many different ways and settings, development of adequate PA skills requires more explicit instruction and training (Lundberg et al., 1988). This need for explicit instruction has implications for how educators target phonological awareness skills and intervene with students who are struggling to develop them.

Learning-Related Behaviors and Academic Achievement

There is much discussion in the literature about how best to target phonological awareness skills and set children on the best trajectory possible. The discussion tends to focus primarily on the instruction and measurement of specific set of early literacy skills and experiences (Hojnoski & Missall, 2007;

National Early Literacy Panel, 2008; National Institute for & National Center for Family, 2008; Phillips, Clancy-Menchetti, & Lonigan, 2008; Whitehurst & Lonigan, 1998). However, a student's level of literacy skill is not the only factor associated with educational success. Preschool children must also demonstrate a set of learning-related behaviors that allow them to adequately participate in a variety of learning activities and environments (McClelland & Morrison, 2003).

Demonstration of these skills can facilitate learning experiences for young children, whereas lack of these positive learning-related behaviors (eg. attention, cooperation) or demonstration of negative behaviors (eg. aggression, anti-social behavior) can impair achievement-oriented behavior (Bierman et al., 2009; DuPaul et al., 2001; McWayne et al., 2004; Normandeau & Guay, 1998).

The relationship is well-established between young children who have difficulty with learning-related behaviors and increased risk for poorer academic outcomes (Arnold, 1997; Buhrmester, Whalen, Henker, MacDonald, & Hinshaw, 1992; Doctoroff, Greer, & Arnold, 2006; DuPaul et al., 2001; Hinshaw, 1992b; McWayne et al., 2004; Normandeau & Guay, 1998; Stage, Abbott, Jenkins, & Berninger, 2003; Vaughn, Hogan, Lancelotta, Shapiro, & Walker, 1992).

McWayne, Fantuzzo, & McDermott (2004) found that certain preschooler classroom competencies (social skills) and approaches to learning (attention, motivation, persistence) were uniquely associated with early academic success for preschool students, even when controlling for demographic and context characteristics. Lonigan et al. (1999) found that behaviors associated with

inattention and hyperactivity were uniquely associated with emergent literacy skills in both low- and middle-income preschool students, and that this relationship was not due to factors associated with general cognitive ability. Doctoroff, Greer, and Arnold (2006) found that for preschool boys, deficits in emergent literacy skills were associated with aggressive behaviors and fewer prosocial interactions.

Researchers have explored a wide range of preschool learning-related behaviors including prosocial peer relations, attention deficits, aggressive behaviors, and self-regulation, as well as differences by sex and age (Bierman et al., 2009; Doctoroff et al., 2006; McClelland et al., 2007; McWayne et al., 2004). Across different types of measurement, including measurement by classroom observation, teacher ratings, and direct assessment, there is agreement that a significant relationship exists in preschool between this category of behaviors and academic outcomes (Bierman et al., 2009; Doctoroff et al., 2006; McClelland et al., 2007; McWayne et al., 2004). Consistently, preschool children who exhibit lower levels of learning-related behaviors, such as failure to demonstrate self-regulation, inhibit aggression, or high levels of inattention are also more likely to have deficits in academic achievement. This relationship has been found to strengthen after children begin and continue through elementary school, particularly in the area of literacy (Miles & Stipek, 2006).

An extensive review of the literature supports the robustness of this relationship for elementary-aged students, but also points to the ambiguity in the

literature surrounding the directionality of the relationship (Hinshaw, 1992b). Do academic deficits cause children to develop behavior challenges? Or do behavioral challenges affect a child's on-task behavior and opportunities to learn? This is still an ongoing debate in both elementary and preschool settings. Dishion (1990) found that elementary-aged children's anti-social behavior led to lower levels of academic achievement, but the reverse relationship – lower academic achievement causing increases in anti-social behavior – was not present. More recently, however, evidence has shown that the relationship between these two constructs in preschoolers and early elementary students represents more of a reciprocal causation model, in that poor academic skills (particularly literacy) lead to behavior challenges and behavior challenges can lead to academic deficits (Arnold, 1997; Trzesniewski, Moffitt, Caspi, Taylor, & Maughan, 2006). Regardless of the directionality of the relationship, it is clear that attention must be paid to both domains in order to fully understand child trajectories and facilitate growth for children with difficulties in these domains.

Defining Learning-Related Behaviors

The literature around academic learning and behavior refers to a variety of similar and interrelated set of behaviors that have shown an association with academic success. However, this set of behaviors is referred to by many different names. These include: Social skills or Pro/Anti-Social Behavior (Miles & Stipek, 2006; Trzesniewski et al., 2006), Behavioral- or Self-regulation (McClelland et al., 2007; Ponitz, McClelland, Matthews, & Morrison, 2009; Skibbe, Connor,

Morrison, & Jewkes, 2011), Attention (Arnold, 1997; Hinshaw, 1992a; Stage et al., 2003), Externalizing Behavior (Arnold, 1997; Hinshaw, 1992b), Classroom or Social Competence (McWayne et al., 2004; Vaughn et al., 1992), and Behavioral or School Readiness (Bierman et al., 2009; La Paro & Pianta, 2000). While there may be conceptual distinctions between these terms, together they encompass a related set of behaviors that children must exhibit or inhibit in order to participate in learning. In addition, all of these behaviors, by any name, have consistently been associated with early academic success.

In an attempt to better understand this broader construct in preschool-aged children, McClelland and Morrison (2003) used confirmatory factor analysis to demonstrate the presence of a construct that combines many of these behaviors in preschool children. McClelland and Morrison (2003, p. 307) refer to this construct as “Learning-related social skills.” It is important to note that this construct is distinct from more general social or interpersonal skills, which might include behaviors like sharing with other children or respecting others. Learning-related social skills represent a specific set of behaviors that are directly related to learning. These behaviors could be conceptualized as the skills necessary to successfully participate in learning-related tasks in the preschool classroom. The behaviors identified by McClelland and Morrison (2003) include:

1. Listening and following directions
2. Participating appropriately in groups (including turn taking)
3. Cooperation

4. Staying on task
5. Organizing work materials
6. Regulation of behavior
7. Self-direction to complete a task

Again, McClelland and Morrison (2003) make a distinction between the above behaviors and behaviors that help children to get along with their peers more generally because of evidence that these are distinct constructs that are differentially related to school performance (McClelland, Morrison, & Holmes, 2000). Their analysis confirmed the presence of the construct of learning-related social skills in preschool students, its stability over time, and the relative stability of preschool student rankings even as skills improved over time (McClelland & Morrison, 2003). Based on their findings, McClelland and Morrison (2003) draw the conclusion that differences in learning-related skills identified in preschoolers as early as age 3 can have important implications for a child's transition to kindergarten and later academic success. Given this, they highlight the importance of schools and teachers having the willingness and ability to adapt expectations and curriculum to suit the differing needs of their students to provide a better match between teacher expectations and level of student skills.

Response to Intervention

One way that educators adapt instruction to the differing needs of students is by providing supplemental or differing intensity of interventions to students who are not making progress in the general curriculum or who come to the classroom

lacking important skills. To develop important early literacy skills, preschool students need instruction or intervention that matches their specific level of need. Response to Intervention (RTI) is one model that has been proposed for use in preschool as a way to provide a match between student skill level and instructional content or intensity of intervention (Greenwood et al., 2011). Response to Intervention (RTI) typically uses a three-tiered model to assess and match students' level of need with the intensity of instruction and intervention they receive (Fuchs & Fuchs, 2006). In a three-tiered model, Tier 1 provides all students with evidence-based, high-quality curriculum, with the expectation that most students will make sufficient progress with this level of support. Tier 2 identifies students—through universal screening—who are not making the expected amount and rate of progress in a certain area and provides them with increased support and more intensive instruction, along with increased frequency of progress monitoring. Tier 3 provides the most individualized, intensive support and instruction to students who continue to show insufficient progress despite Tier 2 supports (Coleman, Roth, & West, 2009). The clear benefit of the RTI model is that it provides an efficient means for educators to screen, monitor, and provide intervention to students who are potentially at-risk before a persistent, long-term pattern of failure emerges.

One of the tenets of RTI is to provide students with evidenced-based interventions that have proven effectiveness. However, even with effective interventions, there are students who do not benefit from these interventions in

the same ways as their peers (Al Otaiba & Fuchs, 2002). In the elementary grades, investigators have reported that as many as 30% of students do not respond to decoding and phonological instruction that is generally effective for the majority of students (Al Otaiba & Fuchs, 2002). Researchers have attempted to identify characteristics of both preschool and elementary students who do not show response to early literacy interventions (Al Otaiba & Fuchs, 2002; Nelson et al., 2003).

A meta-analysis of preschool and elementary learner characteristics that influence early literacy intervention effectiveness found six characteristics that showed moderate effect sizes (Nelson et al., 2003). The three characteristics with the largest effect sizes were (in order of magnitude) rapid naming skills (0.51), problem behavior (0.46), and phonological awareness skills (0.42). The authors highlight that there appears to be very little difference between the contribution of these three characteristics, thus supporting the idea that behavior is an important contributor to a child's responsiveness to early literacy interventions. However, despite recommendations from researchers, some models of RTI do not include consideration or assessment of behavior characteristics in the framework used to place students in to particular academic interventions (Compton et al., 2012). Most RTI systems assess students in academic domains, such as early literacy skills, and use those assessment data to determine what type of intervention a student requires. However, if learning-related behaviors are an important predictor of response to interventions, it

follows that educators may need to assess students in this domain as well within their decision-making framework.

A limited amount of research has examined the effectiveness of including additional factors in the model in order to enhance the accuracy of intervention decision making. For instance, researchers have used teacher ratings of student attention and behavior, in addition to academic screening data, to enhance the ability of decision-making frameworks to predict response to intervention (Compton et al., 2012; Stage et al., 2003). Including teacher ratings of attention and behavior was shown to increase the ability of the models to predict responders and non-responders and therefore their capacity to identify candidates for particular interventions (Compton et al., 2012; Stage et al., 2003). The limited instances of research in this area have included only samples of early elementary students; the functionality of this type of RTI model in preschool has yet to be explored.

Response to Intervention in Preschool

The implementation of RTI in preschool settings is relatively new, therefore many questions remain about how to best implement it with this population (Bayat, Mindes, & Covitt, 2010; C. R. Greenwood et al., 2011; C. R. Greenwood et al., 2008). Similar to the interventions provided under RTI models in the elementary grades, early work on preschool RTI models has developed intervention conditions that provide differing levels of intensity and individual support to students depending on need. In Tier 2 interventions, preschool

students may receive instruction in small groups, with more intensive support than Tier 1, but less individualized support than in Tier 3. For example, one Tier 2 intervention designed for this population is delivered via an interactive, small-group listening center, where students are required to listen and follow along with a guided book activity. (C. R. Greenwood, Carta, Spencer, Guerrero, & Kong, Submitted; Spencer, Goldstein, & Kaminski, 2012). In comparison, a Tier 3 intervention might take the form of one-on-one sessions with a teacher or assistant, where the student receives direct instruction tailored to their ability and where instruction does not progress until the student has mastered the content (Kaminski, Ziolkowski, & Goldstein, 2009).

Logical analysis would suggest that the different intervention conditions require different learning-related skills from students in order for them to participate and learn, or that the presence (or absence) of certain learning-related skills may affect the likely efficacy of a particular form of intervention for some children. Tier 1 interventions, implemented in larger groups, may require more independence from individual students, while Tier 3 interventions implemented in a one-on-one format likely require much less independence from the student. Students demonstrating a lack of positive learning-related behaviors may struggle to learn in a Tier 1 intervention condition, where they are required to demonstrate more independence and rule following, and where teachers may have less ability to provide one-on-one redirection to individual students. Take for example, a listening center intervention developed to provide vocabulary,

phonological awareness, and comprehension instruction (Kaminski et al., 2009). In order to participate in the shared book-reading listening center, students must be able to sit at a table for 10-15 minutes, with a small group of students without being disruptive. In addition, after a brief training period, they must show the ability to wait to open the book until instructed, turn pages only when instructed, and follow directions from the audio recording such as “Point at...”, “Lift the flap”, and or “Thumbs up!” An example of the audio text from this intervention follows:

Do you have your book? Make sure it is closed. Look at the front cover of the book. Now...Do not open it until I tell you to..... Your book should have a picture of Pablo Porcupine on the cover. Make sure Pablo Porcupine is not upside down. Remember do not open the book yet....Look at the top of the book. Do you see the words? This is the title or the name of the book. Do you see the finger with bump on it? touch the bump under the first word. Follow the arrow under the words as we read the title together, it says Pablo's Prickly Problem.

Students demonstrating insufficient learning-related skills as specified by the McClelland and Morrison (2003) model—following directions, independence in a group, cooperation, organizing work materials, and staying on task—may not have the skills to benefit from this type of Tier 2 intervention even if their early literacy skills indicated that Tier 2 was the most appropriate intervention condition. Or put another way, these students’ learning-related skills may not be a good match for the demands of this type of intervention.

Unfortunately, the mismatch of skills to intervention may not be detected by models of RTI designed for early childhood settings that focus solely on language and literacy-related assessment, or it may be assessed separately neglecting the interplay between learning-related behaviors and early academic

success (Barnett et al., 2006). Even researchers in the area of preschool RTI who recognize the importance of attention, engagement, independence, group participation, and the reciprocal relationship between delays in reading and behavior problems have not yet begun to incorporate assessment of these skills into decision-making frameworks that assign students to interventions (C.R. Greenwood et al., 2011). Given the contribution that student behavior has shown in responsiveness to intervention (Nelson et al., 2003), and the wide-spread agreement around the interplay between literacy achievement and behavior, it follows that learning-related behavior should be assessed as part of an RTI model for early childhood. However, the question of how best to assess this domain for the purpose of RTI decision-making remains.

A Decision-Making Framework for Early Childhood RTI

In RTI models, assessment in one domain can be used to determine student need for a particular tier of intervention. For instance, elementary schools may use Curriculum-based Measures (CBMs) of reading fluency to determine need for Tier 2 or 3 instruction in reading (Fuchs & Fuchs, 2006). However, best practice for educational testing dictates that high-stakes instructional decisions should not be made with only one source of data (AERA, 1999). Some would argue that models that make instructional decisions with only one source of data could run the risk of incorrectly classifying students due to insufficient classification accuracy of the measures used to make decisions. The consequences of such errors are that resources may be allocated to students

who are not in need of additional intervention (low sensitivity, or false positives), or that students in need of additional support may not be identified as such (low specificity, or false negatives). Additionally, research on the impact of student behavior on learning indicates that learning-related behavior may need to be included in models used to determine instructional need (Nelson et al., 2003).

One way to increase classification accuracy may be to collect additional sources of data to help enhance the decision-making process (VanDerHeyden, 2011). VanDerHeyden (2011) recommends implementing a gated model of decision-making for RTI, which includes assessment of a variety risk factors (in addition to academic data) to determine need for intervention.

With this concern in mind, The Center for Response to Intervention in Early Childhood (CRTIEC) examined the sensitivity and specificity of its primary measures of early literacy using teacher judgment of tier placement as criterion. The five Individual Growth and Development Indicators (IGDIs) include: Picture Naming, Rhyming, First Sounds, Sound Identification, and Comprehension (McConnell, Wackerle-Holman, & Bradfield, 2012). The sensitivity of IGDIs ranged from .70 to .85; specificity ranged from .46 to .87. In the case of the phonological awareness measures (Rhyming and First Sounds) sensitivity was .71 and .85 respectively; specificity was .70 and .77. If the Rhyming measure alone was used in a RTI model to determine need for Tier 2 or 3 intervention, specificity of .70 indicates that 30% of the students in need of additional PA intervention would not be identified by the measure. This is potentially a

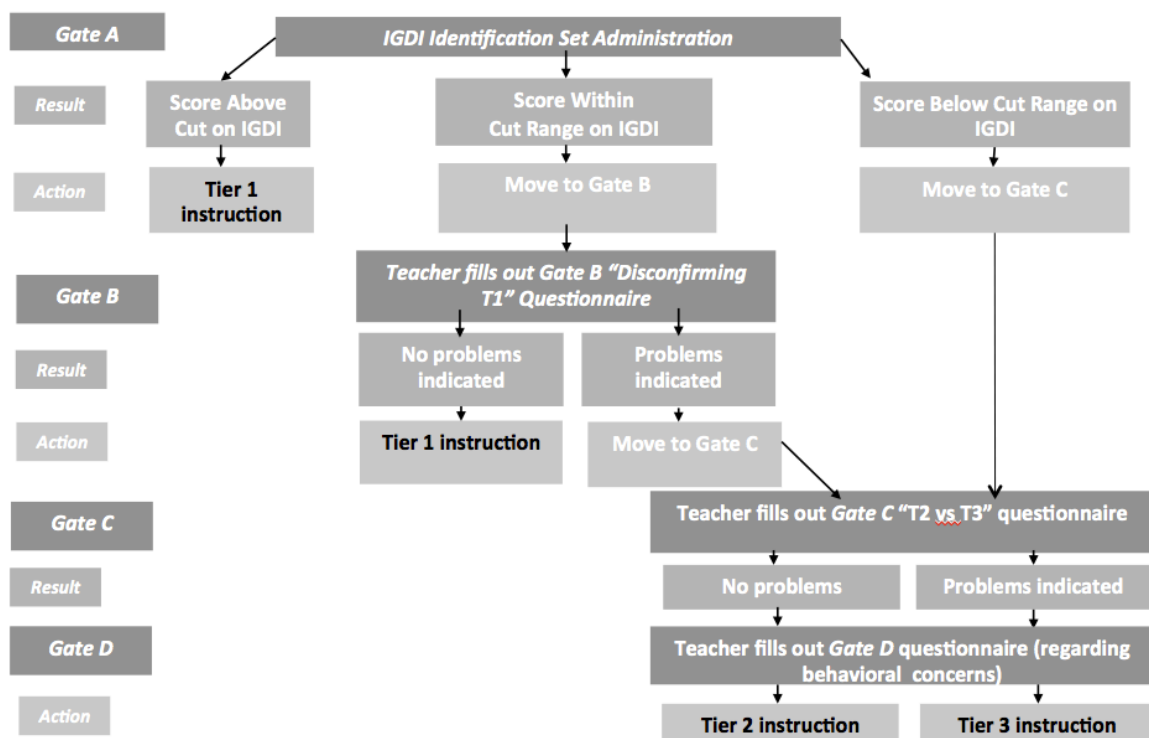
significant number of students who might continue to fail in the general curriculum.

Researchers within CRTIEC have attempted to address this problem by proposing a multiple-gating model of instructional decision-making, which uses multiple sources of information to successively narrow the field of students in need of Tier 2 and Tier 3 interventions (McConnell et al., 2012). The CRTIEC decision-making framework (DMF) includes information from direct assessment of student early literacy skills (IGDIs), as well as teacher ratings of student early literacy skills and learning-related behaviors (see Figure 1).

The DMF consists of four gates through which students are progressively sorted depending on the results of each stage of screening. At Gate A, students are sorted into three groups depending on the results of their IGDI scores. Due to some imprecision in the cut score, the cut score that determines tier placement is actually a range of scores. Tier placement decision is a function of scoring above, within, or below the cut-score range. Student who score above the cut-score range, students who are within the cut-score range, and students who are below the cut-score range. Students above the cut-score range receive Tier 1 instruction, while the other two groups go on to have more information collected about their skills. At Gate B, students who scored within the cut-score range at Gate A are then screened using a teacher questionnaire that gathers more information about their early literacy skills to determine if they have the skills to continue achieving with the instruction and intervention provided in Tier 1.

Students who pass this level of screening receive Tier 1 instruction. All other students proceed to Gate C for further screening. At Gate C, students who scored below the cut-score range at Gate A and students who did not pass screening at Gate B are screened with a teacher questionnaire that is used to determine if they require Tier 3 instruction or if they have the skills to succeed within Tier 2. Before being assigned to Tier 2 or 3 instruction, both of the remaining groups also continue to Gate D, where they are screened with a final teacher questionnaire concerning learning-related behaviors. If no behavior concerns are indicated, the student receives Tier 2 instruction. If significant behavior concerns are indicated, the student either receives Tier 3 instruction or Tier 2 instruction with behavioral supports.

Figure 1. CRTIEC Decision-Making Framework.



For the CRTIEC DMF model to be successful, the decision-making process requires a set of measures designed specifically for the purpose of selecting candidates for intervention. These measures would assess all of the domains that are relevant to predicting learning, including early literacy skills and learning-related behaviors. These measures must be well-suited for the constraints of an RTI model.

Assessing Learning-related Behaviors in a Response to Intervention Model

In order to demonstrate utility in a RTI model, measures must meet certain criteria that are specific to the demands of the model. Among other characteristics, measures must be psychometrically robust enough to make individual-level decisions about students who require more intensive intervention, and they must be logistically feasible (i.e. brief and easy to administer) in order for early educators to easily use with them many students (C. R. Greenwood et al., 2008; Wackerle-Hollman et al., 2013). More specifically, measures of learning-related behavior for an RTI model should be designed for the express purpose of identifying particular learning-related behaviors which enhance or impede learning, and not necessarily for the purpose of diagnostic classification of specific conditions (Barnett et al., 2006). This is a significant distinction given that some children may have behavioral challenges that impact their learning but are not significant enough for formal diagnosis (Arnold, 1997). Measures designed specifically for this purpose would feature assessment of the behaviors or skills most closely related to early academic success (McClelland & Morrison,

2003), rather than a broad or more comprehensive assessment of overall child behavior.

Given the above characteristics, a review of currently available measures of preschool child behavior reveals a lack of appropriate measures for this specific purpose. A variety of published, well-regarded, standardized rating scales of preschool student behavior exist. These include the *Social Skills Rating Scale* (F. M. Gresham & Elliot, 1990), the *Child Behavior Checklist* (T. M. Achenbach & Rescorla, 2000), the *Behavior Assessment System for Children* (Reynolds & Kamphaus, 2010), the *Conners Early Childhood* (Conners & Goldstein, 2009), and the *Preschool and Kindergarten Behavior Scales* (Merrell, 2002). In addition, the literature features other measures of preschool student behavior that are not commercially available, but do have some empirical support, such as the *Adjustment Scales for Preschool Intervention* (Bulotsky-Shearer & Fantuzzo, 2004; Lutz, Fantuzzo, & McDermott, 2002), and the *BASC Teacher Rating Scale for Preschool-Screener* (Distefano & Kamphaus, 2007). While technically adequate, all of these measures are lengthy and time-consuming to complete. They range from 10-30 minutes in estimated administration time, making them impractical for use with larger numbers of students. In addition these measures assess a range of behaviors overly broad in scope for use within an RTI model, with some emphasizing the identification of specific disorders rather than prevention of academic difficulties.

Another method of assessing preschool student behavior included in the literature is direct observation and coding of individual student behaviors. In comparison to rating scales, however, this method is also not very practical for use within an RTI model as it requires significant levels of training to establish agreement on the various behaviors to be coded (eg. eight hours per week for eight weeks), and may require additional staff to conduct the observations (Doctoroff et al., 2006).

While the above-mentioned teacher ratings are too lengthy and broad for use in an RTI model, teacher ratings of learning-related behavior have been found to be a useful component for identifying early learning candidates for intervention and for increasing the accuracy of instructional decision-making (Speece et al., 2011; Stage et al., 2003; Taylor, Anselmo, Foreman, Schatschneider, & Angelopoulos, 2000; Torgesen et al., 1999). One study of teacher ratings of this kind found that beginning kindergarten students identified as at-risk (academically and behaviorally) by their teachers performed more poorly during the school year than a matched group of peers on phonological awareness tasks and had lower ratings of behavior and attention. These deficits continued through the students' first grade year (Taylor et al., 2000). Combining direct assessment of early literacy skills with teacher ratings of learning-related behavior has been shown to enhance prediction of growth in phonological skills (Torgesen et al., 1999). However, for teacher ratings to be useful in an RTI

model, they need to conform more closely to the constraints of the model than do the currently available measures.

Qualities of a Measure of Learning-Related Behavior for Early Childhood

RTI

In an RTI decision-making framework, we are first and foremost concerned with making good decisions for students in order to match them to the intervention condition that will best meet their needs. With any measure that we use in this model, we are concerned with the degree to which it adds some value to the decisions we make about students. In order to assess whether a measure adds value we need certain types of evidence so that we can have confidence in the information we are collecting. In the context of an RTI model, it is necessary to have a measure that both meets the needs of the model (i.e. brief and easy to administer) and possesses certain psychometric characteristics that allow us to have this level of confidence in the inferences we make from its scores. Such a measure should demonstrate precision or consistency in scores and an ability to provide information that is relevant to the inferences to be made from the scores it produces (Thorndike, 2005). These qualities are evaluated by examining the available reliability and validity evidence for a measure.

Reliability Before a measure can add value to a decision-making framework, we must first be able to trust that the scores it generates are sufficiently consistent or reliable. Reliability statistics help us to think about how precisely our assessment measures a given construct, or conversely how much

error in measurement affects the scores that are generated (Cronbach & Shavelson, 2004). Without this consistency, the measure lacks a certain trustworthiness that is necessary when making important decisions for students. There are several ways to examine the reliability of a measure as a function of consistency, including consistency across items as exchangeable indicators of a common construct, the consistency of scores over a period of time, and the consistency of scores over different raters. These factors are important for us to consider if we are to have confidence in the scores a measure produces.

The amount of observed score variance that can be attributed to true score variance, rather than error, is referred to as the measure's internal consistency, and can be assessed using coefficient alpha. If a measure has too much random error included in its scores, the precision of its scores is impacted and we cannot be as confident in the inferences made using these scores. This is an important consideration in an RTI decision-making framework, in that the observed scores on particular measures are used to place individual students in particular intervention conditions. If the observed score contains too much error, educators may make poor decisions using the score and the student may not be placed in the appropriate intervention condition.

When measuring students in the classroom, we may not necessarily be concerned with their particular characteristics at one specific point in time, but rather an overall snapshot of their ability or behavior from day to day. This is particularly true in an RTI model, when we are making decisions about

placement in an intervention condition that will take place over a period time. Measuring in this way allows us to make predictions about what a student's behavior or performance will likely be in the coming days or weeks (Thorndike, 2005). This type of reliability is referred to as test-retest reliability and can be assessed by collecting the same measure at two different time points to demonstrate consistency in scores over time.

Finally, given that scores generated from a rating scale are in essence the judgments of one rater, we are concerned with the generalizability of those scores. It is important for a measure to be able to provide evidence that, to the extent possible, the scores obtained are not the "idiosyncratic results of one rater's subjective judgment" (Tinsley & Weiss, 1975, p. 359), but rather would be similar to scores obtained from other raters. To collect this type of evidence, researchers present information about interrater reliability.

Collecting and assessing all of these types of reliability evidence are important to our understanding of the consistency of our measurement and helps us to gauge the confidence we can place in the scores we collect. Adequate reliability evidence can be thought of as a necessary precondition to establishing the validity evidence of a measure. Once we have information about the trustworthiness of a measure's scores, we can then make a further evaluation of the validity evidence.

Validity Validity evidence helps us make a judgment about the meaningfulness of scores and the appropriateness of using scores for a

particular purpose. Before we use scores gathered from a measure of student characteristics, we need evidence that doing so will add value to the decisions or inferences we make about students. “Validity is an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment” (Messick, 1995, p. 741). In practice, test manuals often present several different types of validity evidence separately (content validity evidence, criterion validity evidence, etc). However, it may be more appropriate to think of validity as a unitary concept, an integration of many types of evidence and an evaluation of them collectively (AERA, 1999). To evaluate validity evidence and know that we are making the best decisions we can, we must ask: to what degree does the entire collection of evidence support the inferences we intend to make using these scores? Is there meaningfulness in the information we are collecting?

To make such a judgment, there must first be a clear statement of the proposed interpretation and use(s) of the measure’s scores (Kane, 2006). This interpretation is linked directly to the construct to be measured, which should be clearly described in terms of its scope and the different aspects of the construct that are to be included in the measure (AERA, 1999). The description of the construct provides a conceptual framework for the measure, which in turn helps to determine what types of validity evidence are necessary to evaluate the inferences made using the scores generated by the measure. In the case of an

RTI framework, our purpose is to use the best quality data to inform decisions about what type of intervention condition best suits the needs of the student.

When collecting data we are attempting to capture a relevant sample of behavior from a larger universe of observations that represent a particular trait. The challenge is to select items or a particular set of observations that include a smaller but representative sample of the larger trait. Evidence should be provided that one can make a generalization from the observed score to what can be expected from a student's performance in a larger universe of observations (Kane, 2006).

Depending on the intended interpretation of scores, it may be desirable to collect evidence that a measure not only captures a representative sample of the larger trait, but also that score variability is not overly influenced by extraneous factors ("skill irrelevant sources of variability") that might influence our interpretation of scores (Kane, 2006, p.24). Collecting this evidence helps avoid the threat to validity known as construct irrelevant variance, or information inadvertently collected by a measure that is irrelevant to the construct of interest (Messick, 1989). To provide this type of evidence, developers should demonstrate that an effort has been made to identify other factors that might produce variability in scores that are not related to the trait of concern. Evidence should also be provided that demonstrates a relationship between the observed scores and other scores associated with the target construct (often demonstrated by a correlation with a criterion measure). This type of evidence is known as

convergent evidence, in that measures of similar constructs would be expected to correlate well with one another. Whereas a weaker relationship between measures of different constructs provides discriminant evidence (AERA, 1999). The above validity evidence demonstrates that a measure collects enough information about a construct to adequately represent it and support inferences made from its scores, but is not collecting information beyond the boundaries of the construct thus weakening support for the inferences being made.

Finally, the administration of a measure is often done with the expectation that some benefit or positive consequence will result from the use of its scores. One of the purposes of the validation process is to determine the likelihood of these benefits being realized, or that the intended consequences of testing will be the actual consequences (AERA, 1999; Messick, 1989). The potential for unintended consequences should be weighed in comparison to the potential benefits of using the measure.

To summarize, when evaluating the validity evidence for a measure we look for the evidence to demonstrate that there is a clearly defined construct to be measured, that the construct is being adequately captured by the measure, that the measure isn't collecting additional irrelevant information, and finally that it is plausible to use the measure to make decisions that will benefit students. Finally, while the concepts of reliability and validity have been described here separately, it is important to recall that reliability is a part of establishing adequate validity evidence. Reliability adds to the validity evidence by describing the

accuracy or sharpness of the tool, whereas the other types of validity evidence provide information about the meaningfulness of using a tool's scores for a particular purpose.

Critique of the Literature

Research clearly demonstrates the importance of building a strong foundation of early literacy skills and the implications for students who fail to develop them (National Early Literacy Panel, 2008; Scarborough, 2001; Scarborough et al., 1991; J. Torgesen et al., 2001; J. K. Torgesen, 2002; Whitehurst & Lonigan, 1998). A clear relationship exists between academic achievement and learning-related behaviors, and more specifically in the area of early academic success and early literacy (Arnold, 1997; Bulotsky-Shearer, Dominguez, Bell, Rouse, & Fantuzzo, 2010; Bulotsky-Shearer & Fantuzzo, 2011; McClelland et al., 2007; Miles & Stipek, 2006; Ponitz et al., 2009). Questions about the directionality of this relationship remain, with some research pointing to a reciprocal relationship and the need for early educators to attend to both academic and behavioral skills (Arnold, 1997; Trzesniewski et al., 2006). There is evidence, however, that behavior has a moderate effect on young students' responsiveness to literacy interventions, comparable to the effect of rapid naming and phonological awareness skills (Nelson et al., 2003). Therefore models of intervention such as RTI should incorporate assessment of certain learning-related behavior into decision-making processes in order to account for the importance of these skills in learning.

More research is needed to help understand the nature of the relationship between learning-related behaviors and the development of early literacy skills. Specifically, questions remain about how best to capture these skills in the context of an early childhood RTI models. Early childhood RTI is still an emerging field and research has yet to look directly at the impact of including behavior in decision-making models for intervention assignment for this age group. In addition, there is currently a lack of measures of learning-related behavior that are appropriate and feasible for use in early childhood RTI systems. Current measures of preschool behavior are overly lengthy and not designed for the purpose of identifying students who have deficits in the skills needed to succeed in particular intervention conditions.

The goal of the present study was to evaluate a newly-developed measure of preschool learning-related behaviors, designed for the purpose of identifying the most appropriate tier of intervention for early literacy instruction in part as a function of behavior characteristics of the student. Reliability and validity evidence were examined in this study to make an overall evaluation of the support for including the measure as a part of an early childhood RTI decision-making framework.

Research Questions

1. What is the extent of the reliability evidence to support the CRTIEC Teacher Questionnaire (CTQ) as a consistent measure of preschool student learning-related behaviors?

- a. What is the level of internal consistency of measurement for the CTQ?
 - b. How consistent are the ratings that teachers provide for the same child at two different times points?
 - c. How consistent are ratings provided by two different teachers for the same student?
2. What is the extent of the validity evidence to support the use of the CRTIEC Teacher Questionnaire (CTQ) as a measure of preschool learning-related behaviors to help inform decisions about tier assignment?
- a. What is the evidence that the CTQ sufficiently captures learning-related behaviors as demonstrated by a correlation between the CTQ and an already established measure of a similar domain?
 - b. What is the extent of the convergent and discriminant evidence to indicate that the CTQ has the capacity to capture learning-related behaviors rather than constructs that it does not purport to measure, as measured by correlations between the CTQ and measures of Phonological Awareness?
 - c. What is the evidence of the CTQ's capacity to add value to an early childhood RTI model by incorporating information about student learning-related behaviors in order to meaningfully influence tier assignment?

Chapter 3

Methods

Participants and Setting

The participants in this study were collected in two different samples, Sample A and Sample B. Sample A was part of a year-long study conducted by the Center for Response to Intervention in Early Childhood (CRTIEC), with data included here collected in the late winter and early spring of 2013. Sample B was a separately collected by the investigator in the spring of the same year. To be included in either sample, students had to be in their P4 year of preschool (entering kindergarten the following fall) and be either 4 or 5 years of age. Teachers in Sample B had to have at least one year of experience and spend at least half of their work day in the classroom in which they were rating students.

Sample A. Sample A was comprised of 295 students, from five school districts in suburban areas surrounding an Upper Midwestern city. All participating sites were part of a larger CRTIEC study already in progress, and the procedures for this study were included in to the CRTIEC application to the Institutional Review Board (IRB) at the University of Minnesota. A subset of the sample, 175 students, were already completing portions of the core battery of measures as part of the standard protocol within the district. Therefore, parents of these students received passive consent forms to opt out of the study, rather than active consent. The remaining 120 students received a larger battery of measures, including a standardized criterion measure. Parents of those students

completed separate consent forms to include the extra measures. The participating sites from Sample A agreed to use the CRTIEC decision-making framework as part of their preschool RTI model. The information in Table 1 reflects the demographic characteristics of the students receiving the larger battery of measures (N=120), including gender, English-language learner (ELL) status, and whether the student was receiving special education services (as indicated by an Individualized Education Plan or IEP).

Sample B. Sample B was comprised of 61 students from 11 childcare and early education centers in and around the same Upper Midwestern city as Sample A. Directors of Sample B centers were recruited by the investigator and were invited to participate either via email or by phone. The center directors then obtained agreement from classroom teacher to participate in the study. All teachers were given the option to opt out of the study if they did not wish to participate. However, none of the invited teachers declined to participate. Three of the centers are located in urban areas, serving predominately low income families. Eight of the centers are in suburban areas serving a range of income levels. All 11 sites had participated in prior research with CRTIEC during previous phases of investigation. Consent was not collected for individual students, as no identifying information was being collected. Instead, teachers received a passive consent form before taking part in the study, with the option to opt out if they did not wish to participate. Table 1 contains demographic information for Sample B, including gender, ELL status, race, and age.

Table 1 Demographic Information for Samples A and B

	Sample A	Sample B
Gender		
Male	63	34
Female	57	27
ELL		
Yes	24	4
No	94	57
IEP		
Yes	27	--
No	93	--
Race		
Black	--	20
Asian	--	6
White	--	34
Hispanic	--	1
Age		
4 years	--	42
5 years	--	19
Total (N)	295	61
Total Receiving Full Battery (N)	120	NA

Note: Different demographic information was collected in Samples A and B. Cells with no values reflect information that was not collected for that sample.

Measures

Early Literacy-Phonological Awareness Phonological Awareness skills were assessed with two measures, the Phonological Awareness Individual

Growth and Development Indicators (IGDIs 2.0), and Test of Preschool Early Literacy (C. J. Lonigan, Wagner, R. K., Torgesen, J. K., & Rashotte, C. A., 2007).

Individual Growth and Development Indicators. *Rhyming* and *First Sounds* (Wackerle-Hollman et al., Submitted) IGDIs were used with Sample A as a screening measure in the late winter and early spring as part of the CRTIEC decision-making framework to assign students to intervention tiers. Each IGDI measure uses a series of 8.5x5.5 inch cards with high-quality photo images of commonly known objects. IGDIs are designed to be easily administered by teachers or other classroom staff and do not require a specialist for administration. In addition, they are designed to be quick, efficient screening of early literacy skills, which make them well-suited for use in an RTI model. For *Rhyming*, preschool students are asked to indicate (by pointing) which image of 2-3 options rhymes with a target image. For *First Sounds*, preschool students are asked to indicate (by pointing) which image of 2-3 options starts with the same sound as the target image. Screening sets were designed using a Rasch model to target the range of expected preschool student ability at specific time points during the P4 year. Reliability estimates are not yet available for these screening sets, however a similar Rhyming screening set for fall has demonstrated reliability of .90, from a sample of 462 students in the fall of 2012 (CRTIEC). Concurrent correlations with a standardized measure of phonological awareness, the TOPEL, were .49 for Rhyming and .61 for First Sounds (Fall 2011 data, CRTIEC). For this study, raw IGDI scores were used for all analyses.

Test of Preschool Early Literacy. The second measure of phonological awareness is a published, standardized measure, the Test of Preschool Early Literacy (TOPEL) phonological awareness subtest (C. J. Lonigan, Wagner, R. K., Torgesen, J. K., & Rashotte, C. A., 2007). A smaller subset of participants in Sample A (N=120) were given the phonological awareness subtest as a criterion measure of phonological awareness. The phonological awareness subtest includes 27 items targeting various aspects of PA skills, including blending, elision, and rhyming. The TOPEL was standardized on a representative sample of 842 U.S. children and has an internal consistency of .87 for the PA subtest. For this study, raw TOPEL PA subtest scores were used for all analyses. The TOPEL was selected for this study because of its ability to aid in the estimation of base rates for the assignment of students to the different RTI tiers within a given sample of students. Given that it is standardized measure, normed on a large representative sample of students, the TOPEL's standard scores (and associated percentile scores) can be used to estimate approximate base rates of the number of students assigned to each tier within an RTI model.

Teacher Ratings-Behavior and Attention Student learning-related behavior was assessed in two ways.

Behavior Assessment System for Children (2nd edition). Students in Sample B were assessed using a standardized measure of preschool student behavior, the Behavior Assessment System for Children, Second Edition, Teacher Rating Scale, Ages 2-5 (Reynolds & Kamphaus, 2010). The BASC

teacher rating scale consists of 100 items and provides several types of scores: an overall behavior composite, an adaptability composite, and several subscales that capture some of the learning-related skills identified by McClelland and Morrison (2003) including aggression, attention, hyperactivity, social skills, and executive function. Teachers rate students on a scale of frequency using: *Never, Sometimes, Often, Almost Always*. For the purposes of this study, the overall behavior composite was examined.

The BASC was selected as a criterion measure for several reasons. While none of the currently available published measures of preschool student behavior were designed specifically to capture learning-related behaviors in an RTI context, the BASC has several qualities that make it a good candidate as a criterion measure of this construct. First, it has a separate form designed specifically for teachers, rather than a combined teacher/parent form, with items designed to capture student behaviors that teachers observe in the classroom setting. In addition, BASC items are written in observable terms to minimize subjectivity. The BASC is also highly correlated with other measures of preschool behavior such as the Achenbach and the Conners. Finally, it provides scores for subscales for scales of interest such as attention, hyperactivity, and aggression that reliably identify populations of students with learning problems such as ADHD and behavior disorders. The BASC Teacher Rating Scale-Preschool reports internal consistency of .96 for the Overall Behavior Symptoms Index, and .91-.95 for the individual behavior composites (Externalizing Behavior,

Internalizing Behavior, and Adaptive Skills). Test-retest reliability is reported as .92 for the Overall Behavior Symptoms Index. Interrater reliability is reported at .76 for the Overall Behavior Symptoms Index (Reynolds & Kamphaus, 2010). For this study, t-scores for each BASC index were used.

CRTIEC Teacher Questionnaire Students in Samples A and B were rated by their teachers using the CRTIEC Teacher Questionnaire (CTQ). The CTQ is a brief rating scale of student behavior that was developed as part of the CRTIEC decision-making framework, in an effort to build a measure that meets the needs of a preschool RTI model. The purpose of the CTQ is to provide educators with pertinent information about student learning-related behavior in order to assist in making informed instructional decisions about student tier placement in early literacy interventions. This information is to be used in addition to screening data from phonological awareness measures in the decision-making framework, as research indicates that both academic and behavior characteristics are important indicators of responsiveness to intervention (Nelson et al., 2003) The CTQ was designed specifically to assess behaviors and student characteristics that might play a role in student success in a particular intervention condition (Tiers 2 or 3). It was designed to identify students that may need additional behavior support or more intensive intervention during phonological awareness instruction due to certain student characteristics, such as ability to independently attend to a small-group activity or the amount of redirection required to keep a student on task. The CTQ consists of 12 items, to which teachers respond either yes or no

depending on the presence or absence of certain characteristics in the student they are rating.

The CTQ was developed as part of a larger decision-making framework for use in an RTI model in preschool. Items were developed in consultation with content area experts and were reviewed by the measurement team of CRTIEC. Content experts and the CRTIEC measurement team reviewed each item for clarity and indicated whether or not the items captured skills that related to a child's ability to participate in each tier of instruction. After reviewing the feedback on the initial pool of items, a set of 20 items were selected for further investigation. This original version of the CTQ was piloted with 24 students from around the Twin Cities metropolitan area. Initial analysis showed internal consistency of .93. After a process of item analysis and removal of poorly functioning items, the internal consistency was .94. The final set of items was further edited for clarity and changed from a 4-point rating scale (ranging from Never/Almost Never to Almost Always/Always) to a yes/no format.

Procedures

Sample A IGDIs were completed on-site at each participating school by classroom staff. Prior to IGDI administration, classroom staff were trained by CRTIEC researchers in standardized procedures for each measure to ensure consistency of administration. Staff were monitored with fidelity checks during administration and received feedback in order to remedy any errors. The TOPEL was administered at each site by trained graduate students with the CRTIEC

project. After each session with the graduate students, the children selected a small toy from a prize box. Teachers of students in Sample A completed the CTQ as part of a larger survey that included items regarding student early literacy and behavior skills (see Appendix A). All students in Sample A had CTQs completed by their teachers. A smaller subset of Sample A also received the IGDIs and TOPEL (N=120).

Sample B CTQ packets were distributed to each of the Sample B sites by the investigator during the first week of the investigation. Packets included a cover sheet that included instructions for selecting students to rate, completing the CTQ, and an offer of a small gift card (\$10) for completing the packet (see Appendix B). Teachers were instructed to complete CTQs for up to four students in their classroom. In the case of classrooms with two teachers or a teacher and an assistant, teachers were provided with instructions to rate the same four students as their co-worker, but to do so separately (in order to collect inter-rater reliability data). Also included in the packets were two BASC protocols, which teachers completed for two of their students: their most challenging student and a student they selected as a typical student for their classroom. Teachers were instructed to complete the packet within a week of when it was distributed to the sites. After one week, the investigator returned to each site to collect the packets and distribute the second packet, which contained a second copy of the CTQ for teachers to complete for the same students during the second week of the investigation, in order to collect test-retest information. After the second week,

the investigator returned to each site to collect the second week's packets and distribute gift cards to participating teachers.

Data Analyses

To answer the first research question regarding the reliability evidence for the CTQ, three different analyses were completed. First, coefficient alpha was calculated to examine the internal consistency of the measure. Next, to examine consistency in teacher ratings between two different time points, test-retest reliability was examined. Finally, to examine consistency between two different teacher ratings for the same student, inter-rater reliability was estimated using the intraclass correlation coefficient.

The second research question, which addressed the extent of the validity evidence to support the use of the CTQ, was answered by examining several types of evidence as well as making an overall evaluative judgment about the evidence as a whole. First, to examine concurrent criterion validity evidence, the correlation coefficient between the scores on the CTQ and a criterion measure, the BASC, was examined. Second, to examine the convergent and discriminant validity evidence for the CTQ, a series of correlation coefficients were examined between the different measures of Phonological Awareness—both the IGDIs and the TOPEL—and the CTQ. Next, to examine the capacity of the CTQ to provide meaningful information in a preschool RTI framework, a series of descriptive analyses were completed to examine the characteristics of the students identified by the CTQ, and how the CTQ's classification of students compares to that of

other measures in the framework. These included an examination of descriptive statistics for the CTQ, and a visual examination of a scatterplot of the CTQ and IGDl scores. In addition, an examination of the classification consistency between the IGDls, the TOPEL, and the CTQ was conducted to look at whether the CTQ provides additional information to the decision-making framework. In other words, do the CTQ, IGDls, and TOPEL scores identify the same students or does the CTQ provide information that may help identify additional students in need of support? As part of this process, a potential cut score for the CTQ was developed and examined. The TOPEL was used in this analysis to estimate base-rates of classification for each of the three RTI tiers. Finally, an overall evaluation of the available validity evidence was conducted to make a comprehensive judgment about the usefulness of the CTQ in the context of a preschool RTI decision-making framework.

CHAPTER 4: RESULTS

This chapter presents the results for each of the research questions posed: (a) internal consistency; (b) test-retest reliability; (c) inter-rater reliability; (d) relation between the CTQ and the BASC (concurrent validity evidence); (e) discriminant validity evidence; (f) capacity of the CTQ to add value to an early childhood RTI model as demonstrated by descriptive analysis the CTQ results.

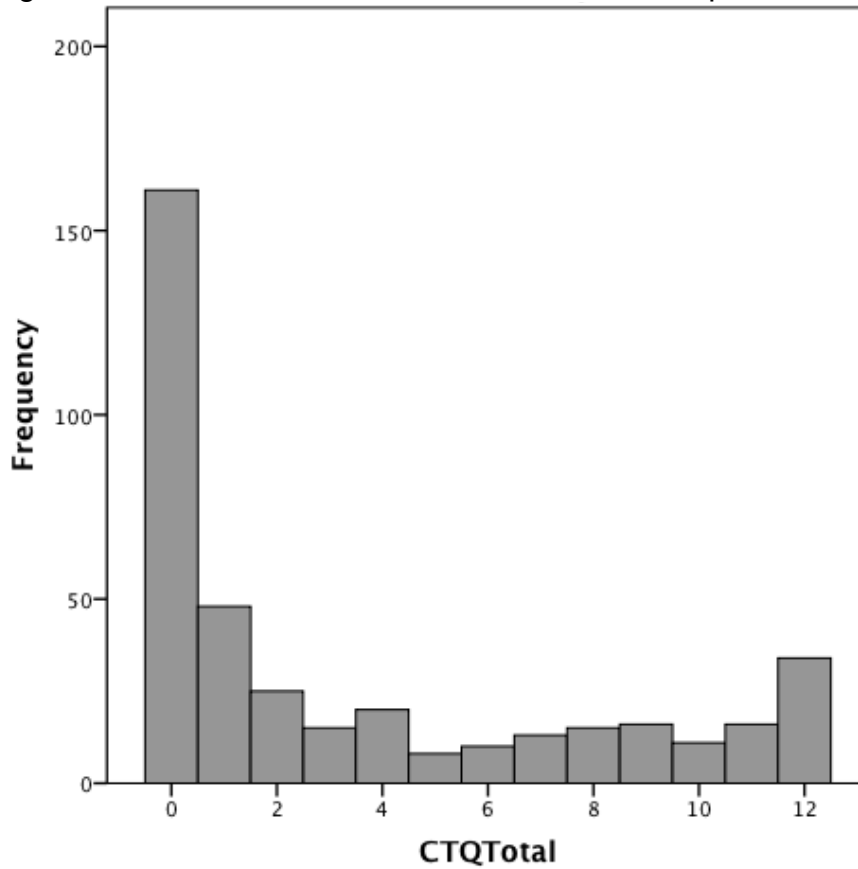
Descriptive Results

Descriptive information for Sample A and Sample B separately, for the sample as a whole, and for each of the criterion measures is presented in Table 2. A visual examination of the CTQ sample as a whole (Figure 2) reveals a bimodal distribution with the most student scores grouped toward the lowest and highest values of the scale.

Table 2. Descriptive Statistics by Measure

Measure	Mean	SD	Skew	Kurtosis	n
CTQ Sample A	3.01	4.21	1.17	-0.20	295
CTQ Sample B	4.88	4.22	.33	-1.30	61
CTQ All Children	3.50	4.29	0.90	-0.74	386
BASC	60.10	15.37	0.14	-1.45	58
TOPEL	91.07	16.06	-0.13	-0.33	120

Figure 2. Distribution of CTQ Total Scores for Samples A and B Combined



In addition to examining descriptive information, and prior to conducting correlational analyses, the assumptions for correlations were examined. This included checking the linearity of the relationship for each correlation, looking for outliers in the scatterplots, and examining the normality of each distribution. As noted, the samples are not normally distributed, implications of this are discussed below. All correlations appeared to have a linear relationship. No significant outliers were observed.

Reliability Analyses

Internal Consistency To answer the research question regarding the internal consistency of the CTQ, coefficient alpha was calculated using data from both Samples A and B, with a total of 386 students. The CTQ scale consisted of 12 items, $\alpha=.92$. It should be noted that alpha's assumption of unidimensionality was not tested in this analysis. In addition, the distribution of CTQ scores is not a normal one, which might also serve to underestimate alpha.

Table 3. Mean, Standard Deviation for CTQ Total Scores

CTQ	Mean	SD	n
All Children	3.5	4.29	386

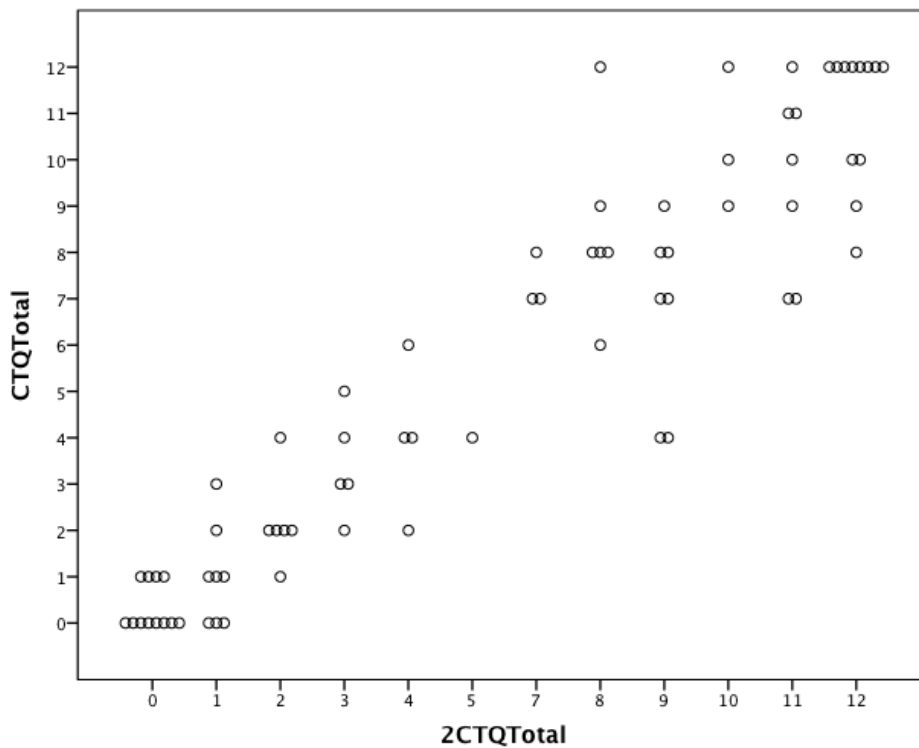
Test-retest Reliability To answer the question regarding the test-retest reliability of the CTQ, a bivariate correlation was calculated between student scores on the CTQ administered at two different time points: Week 1 and Week 2. Seven to ten days elapsed between the first and second administration of the CTQ. Table 4 shows the mean and standard deviation for Week 1 and Week 2 CTQ scores. Sample B data were used to calculate test-retest reliability and included a total of 74 students. CTQ total scores at Week 1 and Week 2 were significantly correlated, $r = .85, p < 0.01$. Visual analysis of the scatterplot of Week 1 and Week 2 scores reveals the presence of what appear to be two distinct scatters of scores. This is not surprising, given that teachers were instructed to rate a typical student and their most challenging student. Again, it should be noted that the distribution of scores was not normal, for the same

reason; two types of students were assessed resulting in scores clustering at the ends of the distribution.

Table 4. Descriptive Statistics for Week 1 and Week 2 CTQ Scores

Scale	Mean	Standard Deviation
Week 1 CTQ Scores	4.99	4.27
Week 2 CTQ Scores	5.93	4.59

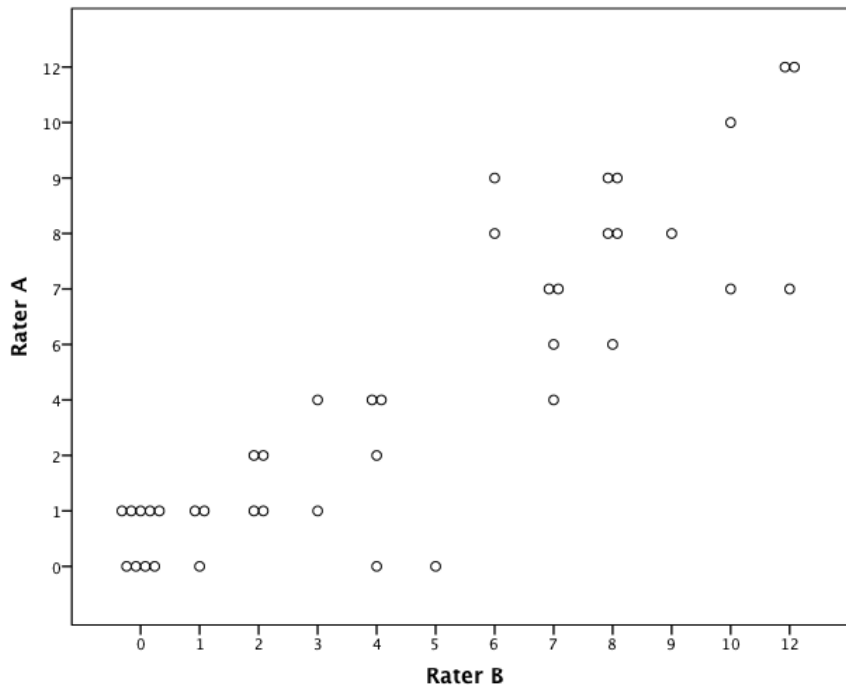
Figure 2. Scatter plot between CTQ scores at Week 1 (CTQTotal) and Week 2 (2CTQTotal)



Inter-rater Reliability To answer the question regarding the inter-rater reliability of the CTQ, an intraclass correlation coefficient was calculated. This calculation helps to estimate the consistency of ratings between two raters who have rated the same student. For this calculation, the one-way random effects model was used, which assumes that the raters are random and the subjects are

fixed. Using data collected from Sample B, with a total of 40 cases from 20 raters, the intraclass correlation for single measures was .89. Again, a visual analysis of the correlation between scores (Figure 3) indicates two distinct scatters of scores, with less representation in the middle of the scale, again likely due to the way the population was sampled (typical students and most challenging students).

Figure 3. Scatter plot between Rater A and Rater B Scores



Validity Evidence Analyses

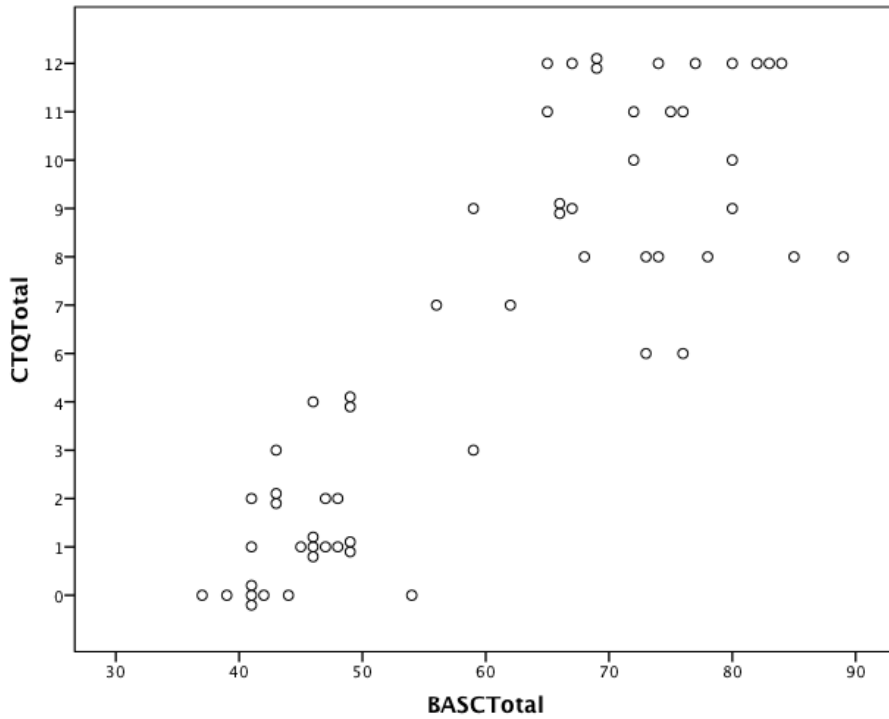
Concurrent Criterion Validity Evidence Concurrent criterion validity evidence was used to examine the extent to which the CTQ measures a similar construct to an already established measure of preschool behavior. For this calculation, student scores from the CTQ were correlated with the T-scores from the overall Behavioral Symptoms Index (BSI) scale from the BASC (Reynolds &

Kamphaus, 2010). Table 5 shows the mean and standard deviation, skew and kurtosis for CTQ and BASC total scores. For the CTQ Total score, the standard deviation is relatively high in comparison to the mean score. In addition, for both measures, the skew is adequate but the kurtosis value for both is a bit high. A Pearson product-moment correlation was used to calculate the relationship between a student's score on the CTQ and their score on the BASC BSI. There was a strong, positive correlation between the two scores ($r = .875, n = 58, p < .01$). Here again, it should be noted that the score distributions were not normal. A visual analysis of the correlation in scores (Figure 4) indicates that the relationship between CTQ and BASC scores was stronger for students at the lower end of the scales (students with more average behavior) and weaker for students at the higher end of the scales (student with more significant behavior concerns).

Table 5. Descriptive Statistics for CTQ Total and BASC Total Scores

Scale	Mean	SD	Skew	Kurtosis
CTQ Total	5.86	4.56	.041	-1.63
BASC Total	60.10	15.37	.136	-1.45

Figure 4. Scatter plot between CTQ Total Scores and BASC Total Scores



Convergent and Discriminant Validity Evidence This portion of the analysis examined whether there was a stronger correlation between like measures than between measures which purport to measure different constructs. For this analysis, Pearson correlations between measures of early literacy (specifically, phonological awareness), between early literacy measures and the CTQ, and between measures of preschool behavior were compared. Raw total scores from the IGDIs Rhyming and First Sounds, the TOPEL Phonological Awareness scale, and CTQ were used. T-scores were used for the BASC. Table 6 contains the correlations between each pair of measures.

Table 6. Correlations between Early Literacy Measures, Behavior Measures, and CTQ.

Measure	IGDI First Sounds (n=120)	TOPEL PA (n=120)	CTQ (n=118)
IGDI Rhyming	.42**	.47**	-.38**
IGDI First Sounds	—	.51**	-.38**
TOPEL PA	—	—	-.42**
BASC BSI	—	—	.88** (n=58)

** p < .01

The magnitude of the correlations between like early literacy measures (TOPEL and IGDI) ranged from between .42-.51. The magnitude of the correlations between like behavior measures (BASC and CTQ) was .88. The magnitude of the correlations between different measures (between TOPEL/IGDI and the CTQ) ranged from .38 to .42. To examine the significance of the difference between each pair of correlations, the Fisher r-to-z transformation was conducted. This transformation transforms r to be normally distributed in order to allow comparisons between the correlations. Table 7 contains the z-scores for the difference between each pair of correlations, as well as the p-values for each comparison. Using the sample size for each correlation, the z-scores were then compared using the formula from Cohen and Cohen (1983). Please note that the absolute value of r was used in each calculation of z, given that we are concerned with the magnitude of the correlation, rather than the observed difference.

Table 7. Testing the Difference in Correlations for Like and Unlike Measures

Correlation 1	Correlation 2	z-score (p)	Type of Comparison
Rhyming/First Sounds (.422)	Rhyming/TOPEL (.472)	-0.48 ($p=.63$)	Like to Like
Rhyming/First Sounds (.422)	First Sounds/TOPEL (.512)	-0.88 ($p=.37$)	Like to Like
Rhyming/First Sounds (.422)	Rhyming/CTQ (-.395)	0.25 ($p=.80$)	Like to Unlike
Rhyming/First Sounds (.422)	First Sounds/CTQ (-.382)	0.36 ($p=.71$)	Like to Unlike
Rhyming/CTQ (-.395)	CTQ/BASC (.875)	-5.71 ($p<.01$)	Unlike to Like
First Sounds/CTQ (-.382)	CTQ/BASC (.875)	-5.81 ($p<.01$)	Unlike to Like
TOPEL/CTQ (-.428)	CTQ/BASC (.875)	-5.47 ($p<.01$)	Unlike to Like

As reported in Table 7, no significant differences for comparisons between correlations of like early literacy measures were observed. Additionally, no significant differences were found in the comparison of the correlations between like IGDl measures and unlike IGDl measures with CTQ. Significant differences were found in the comparison of correlations of unlike measures (early literacy measures with the CTQ) and like behavior measures (the CTQ with the BASC).

Meaningfulness of the CTQ: Additional Analyses

In order to further examine the meaningfulness of the scores from the CTQ, descriptive statistics and a series of descriptive analysis were reviewed. First an examination of descriptive statistics for the CTQ was conducted. Table 8 provides descriptive statistics for the full sample (Samples A and B) as well as descriptive information by Gender and Age for Sample B, gender and age information was not available for Sample A. Review of Table 8 revealed an

apparent significant mean difference in CTQ scores by Gender. To confirm, an independent-samples t-test was conducted to compare CTQ scores by gender. There was a significant difference in CTQ scores for Males ($M=6.62$, $SD=4.36$) and Females ($M=2.89$ $SD=3.52$); $t(59)=3.61$ $p = .001$. No significant difference was observed in CTQ scores by Age.

Table 8. Means, Standard Deviations for Groups by Age and Gender

CTQ	Mean	SD	n
All Children	3.5	4.29	392
4 Years	4.95	4.47	42
5 Years	5.00	4.36	19
Males	6.62	4.36	34
Females	2.89	3.52	27

Next a visual examination of scatterplots of the CTQ and IGDI Rhyming and First Sounds scores, presented in Figures 5 and 6 respectively, was conducted. For this analysis, the Fall IGDI cut scores (and score ranges) developed within the CRTIEC DMF were used as a reference. These score are used in the DMF to identify students as candidates for Tier 2 or 3 instruction. In addition, a preliminary cut score for the CTQ was developed by CRTIEC in consultation with content experts in the area of early childhood RTI. The CRTIEC research team selected a score of 5 on the CTQ as an indicator of potentially significant behavior concerns that may interfere with a student’s learning. The cut scores for both measures are indicated in Figure 5 and Figure 6.

Figure 5 Scatterplot of CTQ total scores and IGD I Rhyming total scores

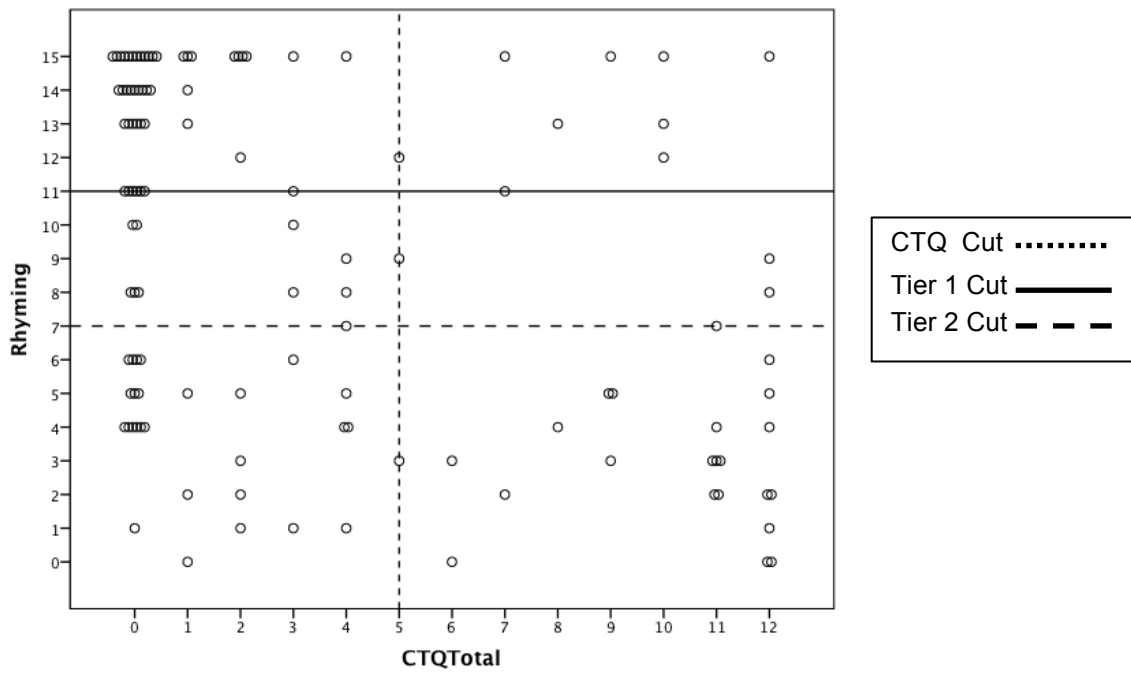
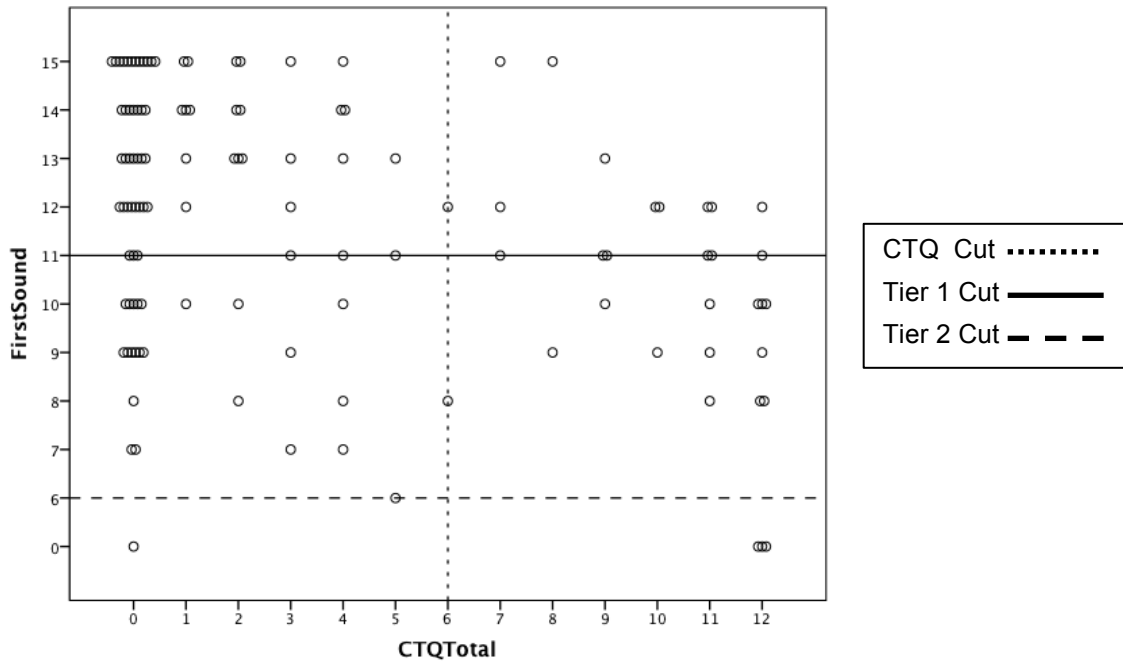


Figure 6 Scatterplot of CTQ total scores and IGD I First Sounds total scores



Finally, an examination of how the IGDIs, CTQ, and TOPEL classify students into tiers within an RTI framework, using data from Sample A was

conducted (Tables 9 and 10). For the purposes of the CRTIEC decision making framework, students were assigned to tiers based on their IGDI scores in Rhyming and First Sounds. If the student was below the cut in *either* Rhyming or First Sounds, they were assigned to the corresponding tier. For the TOPEL, tiers were determined by percentile according to general guidelines for RTI tiers: 1-5 percentile in Tier 3, 6-19th percentile in Tier 2, 20th percentile and higher in Tier 1. For IGDIs+CTQ condition, the preliminary cut score for the CTQ was used to identify students in need of Tier 3 (or extra behavior support). If students in any tier had a CTQ score of 5 or greater, they were moved to the Tier 3 column to indicate a need for additional behavior support. Prior to adding the CTQ to the decision-making criteria, 8 students in Tier 1 and 2 students in Tier 2 (or 8% of the total sample) would have remained in an instructional condition in which their behavior concerns may have interacted with the instructional condition in a way that may have impacted their response to the intervention.

Table 9. Classification of Students into tiers by TOPEL, IGDIs and IGDIs+CTQ (n=120)

Measure	Tier 1 n (percent)	Tier 2 n (percent)	Tier 3 n (percent)
TOPEL	70 (58%)	30 (25%)	20 (17%)
IGDIs	62 (51%)	15 (12.5%)	43 (35.8%)
IGDIs+CTQ	55 (45%)	13 (10.8%)	52 (43.3%)*

*NOTE: Students in this condition had CTQ scores of 5 or greater

To further examine the extent to which the classification of students is consistent across measures, a crosstabs analysis was conducted comparing the classification of students using the CRTIEC decision-making framework with and

without the CTQ. The crosstabs table in Table 10 presents the consistency in tier assignment with the addition of the CTQ to the decision-making framework. Of those students classified as Tier one by the IGDIs, 89% stay in Tier 1 with the addition of the CTQ, while 11% move to Tier 3 based on the CTQ information. Of those students classified as Tier 2 by the IGDIs, 87% stay in Tier 2, while 13% move to Tier 3.

Table 10. Crosstabs Analysis for IGIDIs only and IGDI+CTQ

TierIGDIs * TierIGDICTQ Crosstabulation

			TierIGDICTQ			Total
			1.00	2.00	3.00	
TierIGDIs	1	Count	55	0	7	62
		% within TierIGDIs	88.7%	0.0%	11.3%	100.0%
	2	Count	0	13	2	15
		% within TierIGDIs	0.0%	86.7%	13.3%	100.0%
	3	Count	0	0	43	43
		% within TierIGDIs	0.0%	0.0%	100.0%	100.0%
Total	Count	55	13	52	120	
	% within TierIGDIs	45.8%	10.8%	43.3%	100.0%	

CHAPTER 5: DISCUSSION

The relationship between preschool learning-related behaviors and the acquisition of early literacy skills is well-documented and important for understanding early literacy development (McClelland & Morrison, 2003; McWayne et al., 2004; Nelson et al., 2003). Researchers have shown that learning-related behaviors are uniquely related to early academic success and early literacy acquisition, even when controlling for demographic characteristics, learning context, and cognitive ability (C. J. Lonigan et al., 1999; McWayne et al., 2004). In addition, these behaviors have been found to be as predictive of response to intervention as other learner characteristics such as phonological awareness ability and rapid naming skills (Nelson et al., 2003). Given the importance of learning-related behaviors in predicting early literacy success, it has been suggested that assessment of these skills be included in response to intervention decision-making frameworks (McConnell et al., 2012; VanDerHeyden, 2011). The purpose of the present study was to evaluate a newly-developed measure of preschool learning-related behaviors, designed for the purpose of enhancing identification of the most appropriate tier of intervention for early literacy instruction, in part as a function of behavior characteristics of the student. Reliability and validity evidence was collected and is discussed below, as are the limitations of the present study and implications for future research.

Summary

Results of the present study reveal that the CTQ demonstrated strong internal consistency amongst the items, substantial consistency among ratings across time, and substantial correlation in scores between two independent ratings of the same student. In addition, results of the present study demonstrated that the CTQ measures a construct related to early academic success, was more related to like constructs than unlike constructs, showed an inverse relationship to early literacy skill development, and when used in combination with early literacy measures has the capacity to identify students for intervention differently than with early literacy measures alone.

Reliability

The first set of research questions in this study concerned the collection of reliability evidence for the CTQ. The collection of appropriate reliability evidence is important in order to support specific inferences or score uses. In this case for a measure that may be included in a decision-making framework. This evidence partially supports the confidence with which the measure can be used by demonstrating its consistency or precision.

Internal Consistency The twelve items on the CTQ demonstrated strong internal consistency, indicating that a high degree of score variance observed in the CTQ can be attributed to true score variance. The internal consistency demonstrated by the CTQ surpasses the level of .90 recommended for high stakes decisions. This evidence demonstrates that the CTQ's reliability is

sufficient for making individual-level decisions based on best practice recommendations (Thorndike, 2005). Given the magnitude of reliability observed, the evidence collected here serves to partially support the inclusion of the CTQ in a framework used to make instructional decisions for students in an RTI framework.

Test-Retest Reliability The CTQ administered at two different time points, 7-10 days apart, demonstrated a strong correlation between Week 1 and Week 2 ratings. The agreement between the two time points indicates a high degree of consistency between the ratings, and allows us to infer that teacher ratings of student behavior as measured by the CTQ were relatively stable across the 7-10 day interval. Test-retest reliability evaluates the degree of stability over time for the trait being measured, in this case learning-related behaviors. While preschool student behavior is likely to vary some from day to day or between settings, low consistency (less stability in ratings) over short periods of time could have negative implications for the use of the CTQ as part of an RTI decision-making framework. In an RTI model, students may be in a particular instructional condition for a period of time, at least until the next progress monitoring interval is completed. Therefore, while some day-to-day variability is expected, we look for overall trends in skills and behavior with which to make our decisions. The observed degree of consistency between Week 1 and Week 2 ratings on the CTQ adds to the evidence that the CTQ could potentially be used in an RTI decision-making framework with some confidence.

Inter-rater Reliability One risk in using teacher ratings of behavior to make decisions is their potential for subjectivity (Tinsley & Weiss, 1975). This is a relevant consideration when making decisions about instructional conditions for students, such as those in an RTI framework. Intraclass correlations give us information about the proportion of variance that can be attributed to the target of our measurement, in this case the student, rather than to rater differences (McGraw & Wong, 1996). In this case, we observed a high degree of score consistency between raters, which indicates that that the majority of the variance in scores can be attributed to the student rather than to the rater. The results of this analysis serve to alleviate some of the concern regarding subjectivity of individual raters and further support the consistency of the CTQ and its inclusion in a decision-making framework for preschool students.

Summary of Reliability Evidence Taken together the collection of reliability evidence gathered for the CTQ provides promising evidence about the consistency of the CTQ and adds support for the inclusion of the CTQ in an RTI decision-making framework for preschool students. In the evidence presented here, the CTQ demonstrated high consistency within the measure, high consistency in teacher ratings across a short period of time, and substantial consistency between different teacher's ratings of the same child. While reliability evidence provides only part of the evidence necessary to support the use of the CTQ in an RTI framework, the evidence presented here supports further development of the measure for this purpose.

Validity Evidence

The second set of research questions concerned the collection of validity evidence to support the use of the CTQ as part of an RTI decision-making framework in the context of preschool early literacy instruction. Validity evidence is necessary to support the use of a measure for a particular purpose and gives information about the meaningfulness of a measure's scores. In this case we are looking for evidence that that CTQ has the capacity to provide information that is meaningful to decisions about early literacy instructional conditions for preschool students.

Criterion Validity Evidence The CTQ had a very strong relation to the selected criterion measure of preschool learning-related behavior, the BASC Behavioral Symptoms Index. The BASC was selected as a criterion measure because it is an already established measure of a very similar target domain to that which the CTQ seeks to measure. While these two measures have significant overlap in the target domain they aim to sample, we would not expect a perfect correlation between the BASC and the CTQ. Perfect correlation is not expected because the BASC assesses some aspects of preschooler behavior that the CTQ does not attempt to measure such as atypicality and somatization. However, the amount of overlap in content and the degree of correlation between these two measures provides preliminary evidence that the CTQ is adequately capturing information about preschool learning-related behaviors. Given what we know about the interaction between learning-related behavior and early literacy

acquisition, information about these behaviors is of interest when making decisions in an RTI context. The criterion evidence collected here provides evidence that the CTQ is capturing these important behaviors and adds to the support for the use of the CTQ within an RTI decision-making framework concerning early literacy skill development.

Convergent/Discriminant Evidence The CRTIEC decision-making framework includes both measures of early literacy skills and a measure of behavior. The two types of measures are included in the framework because we propose that they are both potentially relevant when selecting the most ideal early literacy instructional condition for preschoolers, and that decisions might change with the addition of the CTQ. To support the use of the CTQ in the final gate of this decision-making framework, we are looking for evidence that the CTQ is contributing information that is not already collected by the early literacy measures. To collect this type of evidence, the present study collected information about convergent and discriminant validity evidence. Again, convergent validity evidence demonstrates that two measures that we expect to be related, are related. Conversely, discriminant validity evidence supports the expectation that two measures that are not related are indeed not related (AERA, 1999). In this study, our prediction about the relationships between related and unrelated measures is somewhat less succinct. While we expect the measures of like domains (eg. the IGDIs and the TOPEL) to be more highly correlated with one another, we would also expect a lower magnitude but still non-zero

relationship between measures of early literacy and learning-related behaviors because of the interrelatedness of early academic skills and behavior. However, we would predict a weaker correlation between measures of different domains (e.g. CTQ and IGDIs) than we would expect from measures of like domains.

Results of these comparisons indicate that these predictions were primarily supported, but partially not supported. The various measures of early literacy were positively and moderately correlated with one another. The measures of behavior were strongly correlated with one another. Each measure of early literacy was significantly and negatively correlated with the CTQ, which supports the assertion that failure to demonstrate learning related behaviors (as measured by the CTQ) is associated with poorer early literacy skills. Measures of like domains (early literacy or behavior) were significantly and positively correlated with one another.

The remaining question was whether we see a difference in the strength of correlations between like measures and correlations between unlike measures. As expected, we did not see a significant difference when comparing pairs of correlations between like measures, in this case pairs of early literacy measures. However, we expected to see a significant difference in strength of correlation when comparing correlations between like measures and unlike measures to correlations between like measures. This was observed for three of the five comparisons of this type, for the comparisons that included a correlation between an early literacy measure and a behavior measure as compared to a

correlation between like behavior measures. A significant difference was not observed between comparisons that included a correlation between two measures of early literacy compared with a correlation between a measure of early literacy and the CTQ. These results support the relationship between early literacy skills and learning related behavior, while providing some evidence that the CTQ measures a construct not fully captured by the IGDIs.

Capacity to Add Value The final analyses in the present study examined the evidence that CTQ has the capacity to provide information beyond that of the early literacy screening measures alone. An examination of descriptive information for the CTQ indicated that there was a significant difference in mean score for males and females. Given what the literature indicates about the increased frequency and severity of behavior problems in preschool boys as compared to girls, this observed difference is not unexpected (Arnold, 1997; Buhrmester et al., 1992).

The visual examination of scatterplots in Figures 1 and 2 indicates that using early literacy scores alone to determine tier placement would lead to different decisions for some students than if information from the CTQ was included as the final gate of decision-making. Specifically, a number of students (between 10-25% of the sample depending on the measure) in the current study possessed early literacy skills that would indicate compatibility with Tier 1 or Tier 2 instruction, but also demonstrated certain behaviors that may not be compatible with those instructional conditions. It should be noted that the cut

point for the CTQ used in these analyses (a score of 5 or higher) is a provisional one and therefore results should be interpreted with some caution. There is not yet evidence that this score is the point at which the demonstration of these behavior characteristics begins to interact with the appropriate selection of Tier 1, 2, or 3 instruction. This cut point was intended for preliminary exploratory purposes, and would need to be explored further in a subsequent study to confirm its usefulness. This would likely involve the examination of a range of potential cut scores and the outcomes of student assigned to instructional conditions depending on the variation in CTQ cut scores.

The final examination looked at how assignment to particular tiers might vary as a function of the measures used to determine the appropriate instructional condition. Results show that the proportion of children assigned to each tier changed with the addition of the information from the CTQ.

General guidelines for RTI models indicate that the following proportion of students should qualify for each tier: Tier 1: 70-80%, Tier 2: 15-20%, and Tier 3: 5% (Greenwood et al., 2011). It should be noted that these proportions assume a particular population, and may not be appropriate for populations with higher concentrations of at-risk students. In the current study, none of the models and measures used to classify students result in the suggested proportions. All three of the classification models resulted in far more than 5% of students being assigned to Tier 3, with the IGDIs+CTQ model resulting in the most students assigned to Tier 3. Without information about how students would respond in

each condition, it is difficult to determine what percentage of tier assignment is the correct percentage. However, the resulting increased proportion of Tier 3 students has implications for resource allocation and the feasibility of providing Tier 3 instruction to significantly more students than RTI generally anticipates (5%).

Again, the provisional CTQ cut score of 5 was used to inform decision-making in the present study; the change in the number of students assigned to each tier is dependent on this score. The selection of a different CTQ cut score would change the decisions that were made. However, this examination provides preliminary evidence that use of the CTQ changes the decisions made by the CRTIEC decision-making framework for some students. If, as we predict, there is an interaction between the types of behaviors captured by the CTQ and early literacy instruction, changing the decisions made by the DMF based on the demonstration of these behaviors might be appropriate and would be made possible by the inclusion of the CTQ in the framework.

Summary of Validity Evidence In summary, the results presented here provide preliminary evidence that the CTQ has the capacity to provide meaningful information beyond that of early literacy screening measures. There is evidence that the CTQ measures a construct relevant to early academic success: learning-related behavior as demonstrated by the strong correlation between the CTQ and the BASC. In addition, there is preliminary evidence that the CTQ is more related to like constructs (another measure of learning-related

behavior) than to that of unlike constructs (early literacy as measured by the TOPEL and the IGDIs). However, there was still a significant relationship between the CTQ and measures of early literacy, which was not unexpected but also does not provide clear divergent validity evidence. Finally, the use of the CTQ in combination with the IGDIs was shown to identify candidates for the various instructional conditions differently than using IGDIs alone, providing some evidence that the CTQ adds additional information to the decision-making framework. However, this study was not designed in a way that can provide evidence that identifying students in this way leads to improved outcomes.

Limitations

The primary objective of this study was to evaluate the reliability and validity evidence of a new measure of preschool learning-related behavior for potential use in an RTI decision-making framework and to assess whether this measure might add value to identification of students for supplemental instructional support. While some promising evidence was collected in this regard, several aspects of the study design and sample placed limitations on the results and conclusions.

Some of the limitations in the present study include small sample sizes for some of the research questions and descriptive information, a lack of information about any observed differences in functioning of the CTQ by race and special education status, and limitations inherent in the criterion measure to which the

CTQ was compared (the inclusion of domains not assessed by the CTQ and the incorporation of its own degree of error).

This study was also limited in its ability to describe the characteristics of the sample in ways that might assist in explaining the results that were observed for a particular population. For instance, it would be helpful to be able to more fully describe the risk factors present for the students in the study, such as poverty status and more complete information about race and ELL status. Information about these characteristics might influence expectations about the proportion of students assigned to different tiers, in that a higher risk population might expect more students in Tier 2 and Tier 3 than one with few risk factors.

Another limitation of the present study is the preliminary nature of the cut score used for the CTQ in the analyses. While the cut score was selected in consultation with CRTIEC researchers familiar with preschool RTI and learning-related behavior, the selected score lacks an empirical basis at this time. Future studies might examine the appropriateness of this score and evaluate how setting the cut score higher or lower impacts the number and type of students identified as in need of additional behavior support.

Future Directions

One valuable next step for this study would be an examination of student outcomes over time, relative to their scores on the CTQ and other measures. Ideally, subsequent work in this area would examine whether adding the CTQ to the decision-making framework enhances the precision of identification of

students in need of additional support, and in turn enhances outcomes for students in the different tiers of RTI instruction. This would be accomplished by repeating portions of the present study, while adding a second time point with which to collect outcome data.

The proposal to add the CTQ to the decision-making framework is premised on the idea that students lacking certain learning-related behaviors may not respond to interventions in the same way as their peers with similar academic ability. To examine this further, future studies might examine students placed in intervention conditions with and without the input from the CTQ and examine outcomes accordingly. Results of such a study could significantly add to the evidence for the usefulness of the CTQ in an RTI decision-making framework. Without such evidence, the current results are promising but potentially insufficient to support the additional time and resources necessary in adding the CTQ to the DMF framework.

Future studies might also wish to examine teacher perceptions of the CTQ in the context of the larger decision-making framework. Specifically, it would be useful to examine teacher perceptions about the ease of use of the CTQ relative to potential benefits in the classroom. The CRTIEC decision-making framework contains several components, each of which require teacher time and input. It would be useful to gather information about whether teachers perceive the information collected by the CTQ as useful, particularly in enhancing their understanding of the behavioral needs of their students

Conclusions

The present study adds to the evidence for the importance of the connection between preschool early literacy skills and learning-related behaviors. It also provides preliminary evidence of the functionality and usefulness of adding a measure of these learning-related behaviors to a preschool RTI decision-making framework. The present study suggests that a brief, easy to administer measure of preschool learning-related behavior may be capturing information relevant to the success of preschoolers in varying intervention conditions, information that could be incorporated into decision-making protocols to enhance preschool early literacy outcomes.

References

- Achenbach, T. M., & Rescorla, L. (2000). *ASEBA Preschool Forms & Profiles: An Integrated System of Multi-informant Assessment*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education, 23*(5), 300-316.
- Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96*(1), 53-55. doi: 10.1037/0022-0663.96.1.43
- Anthony, J. L., Williams, J. M., McDonald, R., & Francis, D. J. (2007). Phonological processing and emergent literacy in younger and older preschool children. *Annals of Dyslexia, 57*(2), 113-137.
- Arnold, D. H. (1997). Co-occurrence of externalizing behavior problems and emergent academic difficulties in young high-risk boys: A preliminary evaluation of patterns and mechanisms. *Journal of Applied Developmental Psychology, 18*(3), 317-330.

- Barnett, D. W., Elliott, N., Wolsing, L., Bunger, C. E., Haski, H., McKissick, C., & Vander Meer, C. D. (2006). Response to intervention for young children with extremely challenging behaviors: What it might look like. *School Psychology Review, 35*(4), 568-582.
- Bayat, M., Mindes, G., & Covitt, S. (2010). What does RTI (response to intervention) look like in preschool? *Early Childhood Education Journal, 37*(6), 493-500.
- Bierman, K. L., Torres, M. M., Domitrovich, C. E., Welsh, J. A., & Gest, S. D. (2009). Behavioral and cognitive readiness for school: Cross-domain associations for children attending Head Start. *Social Development, 18*(2), 305-323.
- Buhrmester, D., Whalen, C. K., Henker, B., MacDonald, V., & Hinshaw, S. P. (1992). Prosocial behavior in hyperactive boys: Effects of stimulant medication and comparison with normal boys. *Journal of Abnormal Child Psychology, 20*(1), 103-121.
- Bulotsky-Shearer, R. J., Dominguez, X., Bell, E. R., Rouse, H. L., & Fantuzzo, J. W. (2010). Relations between behavior problems in classroom social and learning situations and peer social competence in Head Start and kindergarten. *Journal of Emotional and Behavioral Disorders, 18*(195), 195-210. doi: 10.1177/1063426609351172

- Bulotsky-Shearer, R. J., & Fantuzzo, J. W. (2011). Preschool behavior problems in classroom learning situations and literacy outcomes in kindergarten and first grade. *Early Childhood Research Quarterly, 26*(1), 61-73.
- Bulotsky-Shearer, R., & Fantuzzo, J. (2004). Adjustment scales for preschool intervention: Extending validity and relevance across multiple perspectives. *Psychology in the Schools, 41*(7), 725-736.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*.
- Coleman, M. R., Roth, F. P., & West, T. (2009). Roadmap to pre-K RTI: Applying response to intervention in preschool settings. From National Center for Learning Disabilities, Inc. Retrieved at: <http://www.rtinetwork.org/learn/rti-in-pre-kindergarten>
- Compton, D. L., Fuchs, D., & Fuchs, L. S. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*(2), 394-409.
- Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102*(2), 327-340.
- Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of

- decision rules and procedures. *Journal of Educational Psychology*, 98(2), 394-409.
- Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., Barquero, L., & Bouton, B. (2012). Accelerating chronically unresponsive children to tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of learning disabilities*, 45(3), 204-216.
- Conners, C. K., & Goldstein, S. (2009). *Conners Early Childhood: Manual*. North Tonawanda, NY: Multi- Systems Incorporated.
- Dishion, T. J. (1990). The family ecology of boys' peer relations in middle childhood. *Child Development*, 61(3), 874-892.
- Distefano, C. A., & Kamphaus, R. W. (2007). Development and validation of a behavioral screener for preschool-age children. *Journal of Emotional and Behavioral Disorders*, 15(2), 93-102.
- Doctoroff, G. L., Greer, J. A., & Arnold, D. H. (2006). The relationship between social behavior and emergent literacy among preschool boys and girls. *Journal of Applied Developmental Psychology*, 27(1), 1-13.
- Dunst, C., Trivette, C. M., Masiello, T., Roper, N., & Robyak, A. (2006). Framework for Developing Evidence-Based Early Literacy Learning Practices. *CELLpapers*, 1(1). Retrieved at: http://www.earlyliteracylearning.org/cellpapers/cellpapers_v1_n1.pdf
- DuPaul, G. J., McGoey, K. E., Eckert, T. L., & VanBrakle, J. (2001). Preschool children with attention-deficit/hyperactivity disorder: impairments in

- behavioral, social, and school functioning. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40(5), 508-515.
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41(1), 93-99.
- Greenwood, C. R., Bradfield, T., Kaminski, R., Linas, M., Carta, J. J., & Nylander, D. (2011). The response to intervention (RTI) approach in early childhood. *Focus on Exceptional Children*, 43(9), 1-22.
- Greenwood, C. R., Carta, J. J., Baggett, K., Buzhardt, J., Walker, D., & Terry, B. (2008). Best practices in integrating progress monitoring and response-to-intervention concepts in to early-childhood systems. In J. Grimes & A. Thomas (Eds.), *Best Practices in School Psychology V* (5th ed., Vol. 2, pp. 535-548). Bethesda: National Association of School Psychologist.
- Greenwood, C. R., Carta, J. J., Spencer, E., Guerrero, G., & Kong, N. Y. (Submitted). *The effects of a tier 2 vocabulary and comprehension storybook intervention on preschool children's early learning: A replication*. Center for Response to Intervention in Early Childhood (CRTIEC).
- Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System: Preschool, Elementary Level*. Circle Pines, MN: American Guidance Service.
- Hinshaw, S. P. (1992a). Academic underachievement, attention deficits, and aggression: Comorbidity and implications for intervention. *Journal of Consulting and Clinical Psychology*, 60(6), 893.

- Hinshaw, S. P. (1992b). Externalizing behavior problems and academic underachievement in childhood and adolescence: causal relationships and underlying mechanisms. *Psychological Bulletin*, 111(1), 127.
- Hojnoski, R. L., & Missall, K. N. (2007). Monitoring preschoolers' language and early literacy growth and development. *Young Exceptional Children*, 10(3), 17-27.
- Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (2007). Response to intervention at school: The science and practice of assessment and intervention. *Handbook of Response to Intervention* (pp. 3-9): Springer.
- Kaminski, R., Ziolkowski, R., & Goldstein, H. (2009). *The I in RTI: Effective interventions for language and early literacy*. Paper presented at the 25th Annual Division of Early Childhood (DEC) Conference, Albuquerque, NM.
- Kane, M. T. (2006). Validation. *Educational Measurement*, 4, 17-64.
- La Paro, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research*, 70(4), 443-484.
- Lonigan, C. J., Bloomfield, B. G., Anthony, J. L., Bacon, K. D., Phillips, B. M., & Samwel, C. S. (1999). Relations among emergent literacy skills, behavior problems, and social competence in preschool children from low-and middle-income backgrounds. *Topics in Early Childhood Special Education*, 19(1), 40-53.

- Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). Test of
Preschool Early Literacy: PRO-ED, 8700 Shoal Creek Blvd., Austin, TX.
- Lundberg, I., Frost, J., & Petersen, O.-P. (1988). Effects of an extensive program
for stimulating phonological awareness in preschool children. *Reading
research quarterly*, 263-284.
- Lutz, M. N., Fantuzzo, J., & McDermott, P. (2002). Multidimensional assessment
of emotional and behavioral adjustment problems of low-income preschool
children: Development and initial validation. *Early Childhood Research
Quarterly*, 17(3), 338-355.
- McClelland, M. M., Cameron, C. E., Connor, C., Farris, C. L., Jewkes, A. M., &
Morrison, F. J. (2007). Links between behavioral regulation and
preschoolers' literacy, vocabulary, and math skills. *Developmental
Psychology*, 43(4), 947-959.
- McClelland, M. M., & Morrison, F. J. (2003). The emergence of learning-related
social skills in preschool children. *Early Childhood Research Quarterly*,
18(2), 206-224.
- McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for
early academic problems: The role of learning-related social skills. *Early
Childhood Research Quarterly*, 15(3), 307-329.
- McConnell, S., Wackerle-Holman, A., & Bradfield, T. (2012). *IGDIs and beyond:
Measurement and decision-making for language and literacy RTI efforts in
early childhood education*. Paper presented at the Center for Response to

Intervention in Early Childhood (CRTIEC) 4th Annual Summit on RTI in Early Childhood, Santa Ana Pueblo, New Mexico.

McGraw, K. O., & Wong, S. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30.

McWayne, C. M., Fantuzzo, J. W., & McDermott, P. A. (2004). Preschool competency in context: An investigation of the unique contribution of child competencies to early academic success. *Developmental Psychology*, 40(4), 633-645. doi: 10.1037/0012-1649.40.4.633

Merrell, K. W. (2002). Preschool and Kindergarten Behavior Scales, Second Edition: PRO-ED, 8700 Shoal Creek Blvd., Austin, TX. Web: <http://www.proedinc.com>.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.

Miles, S. B., & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development*, 77(1), 103-117.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading

- development: Evidence From a longitudinal study. *Developmental Psychology*, 40(5), 665-681. doi: 10.1037/0012-1649.40.5.665
- National Early Literacy Panel. (2008). *Developing early literacy report of the National Early Literacy Panel*. Washington, D.C.: National Institute for Literacy.
- Nelson, R. J., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice*, 18(4), 255-267.
- Normandeau, S., & Guay, F. (1998). Preschool behavior and first-grade school achievement: The mediational role of cognitive self-control. *Journal of Educational Psychology*, 90, 111-121.
- Phillips, B. M., Clancy-Menchetti, J., & Lonigan, C. J. (2008). Successful phonological awareness instruction with preschool children: Lessons from the classroom. *Topics in Early Childhood Special Education*, 28(1), 3-17.
- Ponitz, C. C., McClelland, M. M., Matthews, J., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology*, 45(3), 605.
- Reynolds, C. R., & Kamphaus, R. W. (2010). *BASC: Behavior Assessment System for Children* (Second ed.). San Antonio, TX: Pearson.
- Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. B. Neuman &

- D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97-110). New York: Guilford Press.
- Scarborough, H. S., Dobrich, W., & Hager, M. (1991). Preschool literacy experience and later reading achievement. *Journal of Learning Disabilities, 24*(8), 508-511.
- Senechal, M., LeFevre, J., Smith-Chant, B. L., & Colton, K. V. (2001). On refining theoretical models of emergent literacy: The role of empirical evidence. *Journal of School Psychology, 39*(5), 439-460.
- Skibbe, L. E., Connor, C. M., Morrison, F. J., & Jewkes, A. M. (2011). Schooling effects on preschoolers' self-regulation, early literacy, and language growth. *Early Childhood Research Quarterly, 26*(1), 42-49.
- Snow, C. E., Burns, M., & Griffin, P. (1999). Preventing reading difficulties in young children. *Consortium On Reading Excellence, 148-155*.
- Speece, D. L., Schatschneider, C., Silverman, R., Case, L. P., Cooper, D. H., & Jacobs, D. M. (2011). Identification of reading problems in first grade within a response-to-intervention framework. *The Elementary School Journal, 111*(4), 585.
- Spencer, E. J., Goldstein, H., & Kaminski, R. (2012). Teaching vocabulary in storybooks: Embedding explicit vocabulary instruction for young children. *Young Exceptional Children*.
- Stage, S. A., Abbott, R. D., Jenkins, J. R., & Berninger, V. W. (2003). Predicting response to early reading intervention from verbal IQ, reading-related

- language abilities, attention ratings, and verbal IQ—word reading discrepancy: Failure to validate discrepancy method. *Journal of Learning Disabilities*, 36(1), 24-33. doi: 10.1177/00222194030360010401
- Strickland, D. S., & Shanahan, T. (2004). Laying the groundwork for literacy. *Educational Leadership*, 61(6), 74.
- Taylor, H. G., Anselmo, M., Foreman, A. L., Schatschneider, C., & Angelopoulos, J. (2000). Utility of kindergarten teacher judgments in identifying early learning problems. *Journal of Learning Disabilities*, 33(2), 200-210.
- Thorndike, R. M. (2005). *Measurement and Evaluation in Psychology and Education* (Seventh ed.). Columbus, Ohio: Pearson Merrill Prentice Hall.
- Tinsley, H. E., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology*, 22(4), 358-376. doi: 10.1037/h0076640
- Torgesen, J., Rashotte, C., & Alexander, A. (2001). Principles of fluency instruction in reading: Relationships with established empirical outcomes. *Dyslexia, fluency, and the brain*, 333-355.
- Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology*, 40(1), 7-26.
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*, 91(4), 579.

- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*(4), 579-593. doi: 10.1037/0022-0663.91.4.579
- Trzesniewski, K. H., Moffitt, T. E., Caspi, A., Taylor, A., & Maughan, B. (2006). Revisiting the association between reading achievement and antisocial behavior: New evidence of an environmental explanation from a twin study. *Child Development, 77*(1), 72-88.
- VanDerHeyden, A. M. (2011). Technical adequacy of response to intervention decisions. *Exceptional Children, 77*(3), 335-350.
- Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities: Research & Practice, 18*(3), 137-146.
- Vaughn, S., Hogan, A., Lancelotta, G., Shapiro, S., & Walker, J. (1992). Subgroups of children with severe and mild behavior problems: Social competence and reading achievement. *Journal of Clinical Child Psychology, 21*(2), 98-106.
- Wackerle-Hollman, A., Schmitt, B., Bradfield, T., McConnell, S., & Rodriguez, M. (2013). Redefining individual growth and development indicators: Phonological awareness. *Journal of Learning Disabilities, 0022219413510181*.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, 69, 848-872.

Appendix A

CRTIEC Teacher Questionnaire (Sample A Version)

For the following questions, please indicate whether the child does or does not demonstrate the indicated behavior by making a check in the “yes” or “no” box.

Part 3. Child Characteristics	Yes	No
31. This child needs frequent prompting to follow classroom rules (eg. hands to yourself).	<input type="checkbox"/>	<input type="checkbox"/>
32. This child has difficulty staying engaged during small group activities.	<input type="checkbox"/>	<input type="checkbox"/>
33. This child has difficulty staying focused on interactive tasks for 15 minutes.	<input type="checkbox"/>	<input type="checkbox"/>
35. This child requires multiple teaching strategies in order to learn something new (eg. Hearing a story about it, then doing an activity, then practicing).	<input type="checkbox"/>	<input type="checkbox"/>
36. This child is easily distracted by peers in a small group setting.	<input type="checkbox"/>	<input type="checkbox"/>
37. This child requires frequent redirection to remain on task.	<input type="checkbox"/>	<input type="checkbox"/>

	Yes	No
38. This child needs high levels of positive verbal reinforcement to complete a task.	<input type="checkbox"/>	<input type="checkbox"/>
39. This child’s behavior interferes with his or her learning.	<input type="checkbox"/>	<input type="checkbox"/>
40. This child needs prompting to stay on task during a 10-15 minute activity.	<input type="checkbox"/>	<input type="checkbox"/>
41. This child learns best in an individual instruction (one-on-one) format.	<input type="checkbox"/>	<input type="checkbox"/>
42. This child needs assistance understanding basic instructions.	<input type="checkbox"/>	<input type="checkbox"/>
43. This child requires a lot of repetition when learning new skills.	<input type="checkbox"/>	<input type="checkbox"/>

Appendix B

CRTIEC Teacher Questionnaire Packet (Sample B Version)

Survey Instructions for Co-Teachers (or Teachers plus Assistants):

- You will be rating the **SAME 4 STUDENTS** as your coworker.
- Read the descriptions for each type of student on the student info sheets (Student A,B,C,D) and decide **with your coworker** which 4 students you will rate.
- **Write each student's initials** at the top of each student info sheet (Students A, B,C,D). Make sure they are the **SAME** as your coworker.
- Complete:
 - o The student information sheet for each student.
 - o The behavior ratings scales for each of the 4 students.
 - o The BASC rating scales for Student A and Student B (attached)
- Return **BOTH TEACHERS' FULLY-COMPLETED SURVEYS** to the envelope provided and **submit it to your center director**.
- A very brief follow-up survey will be given in one week.

TO RECEIVE YOUR TARGET GIFT CARD PLEASE PROVIDE THE FOLLOWING INFO:

TEACHER NAME: _____
CENTER NAME: _____

(Your name will not be stored with the final survey results, your responses will be kept confidential.)



STUDENT A INFO SHEET:

Please choose the student in your classroom who has the ***most difficulty with behavior in your classroom*** (following directions, getting along with peers, sitting still etc).

Child's Initials _____

Child's Ethnicity (check all that apply):

_____ Black / African-American

_____ Hispanic / Latino

_____ Asian / Asian-American

_____ Native American

_____ White / Caucasian

_____ Other :

Age:

_____ 4 years _____ 5 years (Note: Please no 3 year olds)

Gender:

_____ Male

_____ Female

Is this child an English Language Learner?

_____ English is their first language.

_____ This child is learning English as their second language.

STUDENT B INFO SHEET:

Please choose the student in your classroom who shows ***typical behavior in your classroom*** (shows behavior that is average compared to other kids in your room).

Child's Initials _____

Child's Ethnicity (check all that apply):

_____ Black / African-American

_____ Hispanic / Latino

_____ Asian / Asian-American

_____ Native American

_____ White / Caucasian

_____ Other :

Age:

_____ 4 years _____ 5 years (Note: Please no 3 year olds)

Gender:

_____ Male

_____ Female

Is this child an English Language Learner?

_____ English is their first language.

_____ This child is learning English as their second language.

STUDENT C INFO SHEET:

Please choose **ANY** student in your classroom who you haven't already rated.

Child's Initials _____

Child's Ethnicity (check all that apply):

_____ Black / African-American

_____ Hispanic / Latino

_____ Asian / Asian-American

_____ Native American

_____ White / Caucasian

_____ Other :

Age:

_____ 4 years _____ 5 years (**Note:** Please no 3 year olds)

Gender:

_____ Male

_____ Female

Is this child an English Language Learner?

_____ English is their first language.

_____ This child is learning English as their second language.

STUDENT D INFO SHEET:

Please choose **ANY OTHER** student in your classroom who you haven't already rated.

Child's Initials _____

Child's Ethnicity (check all that apply):

_____ Black / African-American

_____ Hispanic / Latino

_____ Asian / Asian-American

_____ Native American

_____ White / Caucasian

_____ Other :

Age:

_____ 4 years _____ 5 years (**Note:** Please no 3 year olds)

Gender:

_____ Male

_____ Female

Is this child an English Language Learner?

_____ English is their first language.

_____ This child is learning English as their second language.

Student Behavior Questionnaire STUDENT A (Most Difficult Student)

Child's Initials _____

For the following questions, please indicate whether the child does or does not demonstrate the indicated behavior by making a check in the "yes" or "no" box.

	Yes	No
1. This child needs frequent prompting to follow classroom rules (eg. hands to yourself).		
2. This child has difficulty staying engaged during small group activities.		
3. This child has difficulty staying focused on interactive tasks for 15 minutes.		
4. This child requires multiple teaching strategies in order to learn something new (eg. Hearing a story about it, then doing an activity, then practicing).		
5. This child is easily distracted by peers in a small group setting.		
6. This child requires frequent redirection to remain on task.		

	Yes	No
7. This child needs high levels of positive verbal reinforcement to complete a task.		
8. This child's behavior interferes with his or her learning.		
9. This child needs prompting to stay on task during a 10-15 minute activity.		
10. This child learns best in an individual instruction (one-on-one) format.		
11. This child needs assistance understanding basic instructions.		
12. This child requires a lot of repetition when learning new skills.		

Student Behavior Questionnaire STUDENT B (Typical Student)

Child's Initials _____

For the following questions, please indicate whether the child does or does not demonstrate the indicated behavior by making a check in the "yes" or "no" box.

	Yes	No
1. This child needs frequent prompting to follow classroom rules (eg. hands to yourself).		
2. This child has difficulty staying engaged during small group activities.		
3. This child has difficulty staying focused on interactive tasks for 15 minutes.		
4. This child requires multiple teaching strategies in order to learn something new (eg. Hearing a story about it, then doing an activity, then practicing).		
5. This child is easily distracted by peers in a small group setting.		
6. This child requires frequent redirection to remain on task.		

	Yes	No
7. This child needs high levels of positive verbal reinforcement to complete a task.		
8. This child's behavior interferes with his or her learning.		
9. This child needs prompting to stay on task during a 10-15 minute activity.		
10. This child learns best in an individual instruction (one-on-one) format.		
11. This child needs assistance understanding basic instructions.		
12. This child requires a lot of repetition when learning new skills.		

Student Behavior Questionnaire STUDENT C

Child's Initials _____

For the following questions, please indicate whether the child does or does not demonstrate the indicated behavior by making a check in the "yes" or "no" box.

	Yes	No
1. This child needs frequent prompting to follow classroom rules (eg. hands to yourself).		
2. This child has difficulty staying engaged during small group activities.		
3. This child has difficulty staying focused on interactive tasks for 15 minutes.		
4. This child requires multiple teaching strategies in order to learn something new (eg. Hearing a story about it, then doing an activity, then practicing).		
5. This child is easily distracted by peers in a small group setting.		
6. This child requires frequent redirection to remain on task.		

	Yes	No
7. This child needs high levels of positive verbal reinforcement to complete a task.		
8. This child's behavior interferes with his or her learning.		
9. This child needs prompting to stay on task during a 10-15 minute activity.		
10. This child learns best in an individual instruction (one-on-one) format.		
11. This child needs assistance understanding basic instructions.		
12. This child requires a lot of repetition when learning new skills.		

Student Behavior Questionnaire STUDENT D

Child's Initials _____

For the following questions, please indicate whether the child does or does not demonstrate the indicated behavior by making a check in the "yes" or "no" box.

	Yes	No
1. This child needs frequent prompting to follow classroom rules (eg. hands to yourself).		
2. This child has difficulty staying engaged during small group activities.		
3. This child has difficulty staying focused on interactive tasks for 15 minutes.		
4. This child requires multiple teaching strategies in order to learn something new (eg. Hearing a story about it, then doing an activity, then practicing).		
5. This child is easily distracted by peers in a small group setting.		
6. This child requires frequent redirection to remain on task.		

	Yes	No
7. This child needs high levels of positive verbal reinforcement to complete a task.		
8. This child's behavior interferes with his or her learning.		
9. This child needs prompting to stay on task during a 10-15 minute activity.		
10. This child learns best in an individual instruction (one-on-one) format.		
11. This child needs assistance understanding basic instructions.		
12. This child requires a lot of repetition when learning new skills.		