

Southern Methodist University

SMU Scholar

Department of Teaching and Learning Theses
and Dissertations

Department of Teaching and Learning

Summer 2020

Making the Case for the Systematic Observation of Language and Reading: A Reading Comprehension Measure Inclusive of Students with Disabilities

Carlin Conner
carlinc@smu.edu

Follow this and additional works at: https://scholar.smu.edu/simmons_dtl_etds



Part of the [Special Education and Teaching Commons](#)

Recommended Citation

Conner, Carlin, "Making the Case for the Systematic Observation of Language and Reading: A Reading Comprehension Measure Inclusive of Students with Disabilities" (2020). *Department of Teaching and Learning Theses and Dissertations*. 11.

https://scholar.smu.edu/simmons_dtl_etds/11

This Dissertation is brought to you for free and open access by the Department of Teaching and Learning at SMU Scholar. It has been accepted for inclusion in Department of Teaching and Learning Theses and Dissertations by an authorized administrator of SMU Scholar. For more information, please visit <http://digitalrepository.smu.edu>.

Making the Case for the Systematic Observation of Language and Reading: A Reading
Comprehension Measure Inclusive of Students with Disabilities

Carlin Conner

Southern Methodist University

Author Note

Funding for my research, including this dissertation work, was supported by grant
#H325H140001 from the Institute for Education Sciences (IES). This grant supported me as a
fellow in the National Center for Leadership in Intensive Intervention (NCLII) from 2016-2020

Copyright © 2020 Carlin Conner

All rights reserved

Table of Contents

Acknowledgements	4
Abstracts	5
Chapter I: Introduction.....	8
Chapter II: A Synthesis of Reading Comprehension Measures for Students with Autism Spectrum Disorder and Low IQ.....	13
References	38
Tables and Figure	46
Chapter III: Examining the Technical Adequacy of the Systematic Observation of Language and Reading.....	53
References	75
Tables	80
Chapter IV: Exploring the Sensitivity of the Systematic Observation of Language and Reading across Students with Intellectual Disability and Autism Spectrum Disorder	86
References	102
Tables and Figure	106
Chapter V: Conclusion	113
Additional References.....	116
Appendix.....	119

Acknowledgements

My PhD journey would not have been possible without the support of the National Center for Leadership on Intensive Intervention (NCLII). My participation in NCLII provided professional opportunities that have been crucial to my development as an education researcher and scholar. No opportunity has been greater than the opportunity to meet my closest mentors, colleagues, and friends from universities involved with NCLII. Without you all, my experiences as a scholar would have been far less meaningful.

Thank you to the numerous faculty members who have supported me. First, thank you Jill Allor, for all the time you have spent molding me into the scholar that I have become. I am appreciative of the endless time you have spent with me, providing feedback and guidance, and making me feel supported and understood throughout this PhD journey. Thank you to the other members of my dissertation committee, Devin Kearns, Francesca Jones, Amy Rouse, and Paul Yovanoff. Devin, I will take my “kearnels” of knowledge from you throughout all my future endeavors. Amy, thank you for being interested in my work every time you walked by my desk and for all the corona zoom calls. Paul, thank you for constantly opening your door to me, the time I spent learning from you has helped form my research interests over the past few years, and I will always be grateful for you. Francesca Jones and Diane Gifford, thank you for involving me in the development of your observation tool; this dissertation would have never happened without support from the two of you. Stephanie Al Otaiba, thank you for your mentorship, which has influenced me professionally and personally, far more than you will ever know.

Finally, thank you to Kristi, Jennifer, Eva, and Sumei. The past four years have been some of the most challenging and unpredictable years of my life, and there are no other people with whom I would have rather completed this ride. See you all at the next zoom Happy Hour!

Abstracts

Paper I:

A Synthesis of Reading Comprehension Measures for Students with Autism Spectrum Disorder and Low IQ

The purpose of this literature synthesis is to examine the empirical literature on comprehension measures used in studies of reading interventions for students with autism spectrum disorder (ASD) and low IQ. Sixteen studies met inclusion criteria for this review. All studies were experimental, included a measure of reading comprehension, and sampled participants with ASD and low IQ (i.e., at least one standard deviation below the mean). The focus of this synthesis is on the comprehension measures used in the interventions, including type of measure and comprehension construct assessed through the measure. Results indicate that most of the studies included measures of comprehension that the researchers created for their study, while a few researchers relied on published measures. Additionally, among the assessments, five different comprehension behaviors were assessed: passage reading comprehension, supported passage reading comprehension, sentence/phrase comprehension, listening comprehension, and literacy engagement. The measures and the comprehension behaviors they assessed are discussed in detail. Implications for research and practice involving reading comprehension measures to inform research on reading interventions for individuals with ASD and low IQ are also discussed.

Key words: autism spectrum disorder, low IQ, reading comprehension, comprehension measure(s), comprehension behavior(s)

Paper II:

Examining the Technical Adequacy of the Systematic Observation of Language and Reading

This paper describes a researcher-created measure of reading comprehension behaviors that is designed to assess reading comprehension growth exhibited by students with Intellectual Disability (ID) and Autism Spectrum Disorder (ASD), the Systematic Observation of Language and Reading (SOLR), and the techniques used to determine its reliability and validity. The SOLR is an observation tool that contains 7 constructs of behaviors related to reading comprehension: (a) language development, (b) abstract thinking, (c) elaboration, (d) print, (e) engagement, (f) fluency and prosody, and (g) off task/refusal behaviors. The seven constructs are captured by 25 specific behaviors that the rater observes or does not observe, at 30-second intervals during literacy instruction. The SOLR was created in response to a study examining the response of students with ID to a comprehensive literacy intervention, during which researchers found the need to develop a measure of reading comprehension behaviors sensitive enough to capture change in students with intensive needs such as ID, ASD, and low IQ. In this paper, I discuss the methods used to determine interrater reliability of the SOLR, the issues around the best reliability statistic to use in the presence of high rater agreement, and whether or not rater agreement is conditional on behavior construct. I also examine internal validity and development of reading comprehension constructs listed on the tool. Implications for practice and use of the observation tool for practitioners and researchers, including lines of future research, are also discussed.

Key words: rating scales, observation tool, literacy engagement, intellectual disability, autism spectrum disorder, reading comprehension measure, reading comprehension behavior

Paper III:

**Exploring the Sensitivity of the Systematic Observation of Language and Reading across
Students with Intellectual Disability and Autism Spectrum Disorder**

This paper explores the sensitivity of a researcher-created measure of early comprehension behaviors, the Systematic Observation of Language and Reading (SOLR), across students with Intellectual Disability (ID) and Autism Spectrum Disorder (ASD) over time. The SOLR is an observation tool that measures early reading comprehension behaviors, such as language development, abstract thinking, elaboration, print, engagement, fluency/prosody, and off task/refusal behaviors. In this study, the sensitivity of the SOLR is explored by examining the progress of 12 students with ID and ASD in response to a reading intervention designed for students with disabilities. The 12 participants range in age from 6-12 years and have IQs spanning from 40 to 70. The SOLR is designed for use with video data, which was collected from the students during the beginning and end of the intervention. Results from the t-tests used to examine change in score over time across all participants showed that the SOLR constructs were not sensitive to change over time; however, examination of individual student data shows that for some students there was improvement in all comprehension constructs measured by the SOLR. Explanation for the lack of significant findings includes discussion of the video data used in the study that was initially collected for purposes other than capturing comprehension behaviors. Examination of the video data supports that the SOLR likely is sensitive to change, and future research should include a plan for data collection that eliminates the variability contributing to differences in student opportunities to respond across videos.

Key words: observation tool, intellectual disability, autism spectrum disorder, reading comprehension measure, reading comprehension behavior

Chapter I: Introduction

This dissertation addresses gaps in the literature surrounding measurement of reading comprehension behaviors for students with Intellectual Disability (ID) and Autism Spectrum Disorder (ASD) through a three-paper dissertation. Specifically, I focus on the development of an early literacy observation instrument, the Systematic Observation of Language and Reading (SOLR), which includes observation of reading comprehension processes such as literacy engagement, oral language, and comprehension. This three-paper dissertation is comprised of: 1) a review of literature synthesizing studies that include comprehension measures used in interventions for student with ASD and low IQ; 2) a report of the technical adequacy of the Systematic Observation of Language and Reading (SOLR); and 3) an examination of the sensitivity of the SOLR over time with a group of students with ID and ASD in response to a comprehensive literacy intervention.

Paper One

The purpose of paper 1, titled *A Synthesis of Reading Comprehension Measures for Students with Autism Spectrum Disorder and Low IQ*, is to examine the empirical literature on comprehension measures used in studies of reading interventions for students with Autism Spectrum Disorder (ASD) and low IQ. Reading comprehension, an active process that occurs when a reader gathers meaning from text, has been considered by researchers to be one of the most important skills learned in school for all students (Bursuck & Darner, 2011; Chaing & Lin, 2007; Masteropieri & Scruggs, 1997). Reading comprehension is also an important skill for students with disabilities, such as those with ASD and low IQ, as it is required to navigate life successfully outside of the classroom (Mims et al., 2012). In order to assess response of students with ASD and low IQ to reading comprehension interventions, researchers must use accurate and

reliable assessments. However, many reading comprehension assessments have narrow response systems and do not capture other types of behavior necessary to develop reading comprehension, such as oral language, engagement, or listening comprehension (Cutting & Scarborough, 2006; Foorman et al., 2018; Francis et al., 2006).

In paper 1, I explore the measures used to capture comprehension growth of students with ASD and low IQ by examining the types of measures used by researchers in experimental studies of reading comprehension interventions for this population, as well as the comprehension constructs those measures assess. In most of the sixteen studies that met criteria for the synthesis, researchers included measures of comprehension created for their study, while a few researchers relied on published measures. Among the assessments, five different comprehension behaviors were assessed: passage reading comprehension, supported passage reading comprehension, sentence/phrase comprehension, listening comprehension, and literacy engagement. Implications for research and practice involving a reading comprehension measure that captures all of these comprehension behaviors, as well as other related behaviors, is described in the discussion section.

Paper Two

The purpose of paper 2, titled *Examining the Technical Adequacy of the Systematic Observation of Language and Reading*, is to describe a researcher-created measure of reading comprehension behaviors that is designed to assess reading comprehension growth exhibited by students with Intellectual Disability (ID) and Autism Spectrum Disorder (ASD), the Systematic Observation of Language and Reading (SOLR). The need for an assessment such as the SOLR was described in the discussion section of paper 1, and paper 2 focuses on the development of this tool and the techniques used to determine its reliability and validity.

The SOLR is an observation tool that contains 7 constructs of reading comprehension behaviors: (a) language development, (b) abstract thinking, (c) elaboration, (d) print, (e) engagement, (f) fluency and prosody, and (g) off task/refusal behaviors. The seven constructs are captured by 25 specific behaviors that the rater observes, or does not observe, at 30-second intervals during literacy instruction, best used with video footage. In this paper, I discuss the methods used to determine interrater reliability of the SOLR, the issues around the best reliability statistic to use in the presence of high rater agreement, and whether or not interrater reliability is conditional on behavior construct. I also examine internal validity and development of reading comprehension constructs listed on the tool to determine how well the 25 behaviors align to their intended construct of reading comprehension. Implications for practice and use of the SOLR by practitioners and researchers are discussed, including a future line of research in which the sensitivity of the SOLR is examined over time across students with ID and ASD.

Paper Three

The purpose of paper 3, titled *Exploring the Sensitivity of the Systematic Observation of Language and Reading across Students with Intellectual Disability and Autism Spectrum Disorder*, is to examine whether the SOLR is sensitive enough to detect change in comprehension behavior of students with disabilities over time. The SOLR is designed for use with video data, which was collected from 12 students with Intellectual Disability (ID) and Autism Spectrum Disorder (ASD) during participation in the efficacy study of a comprehensive reading intervention. The 12 participants range in age from 6-12 years and have IQs spanning from 40 to 70. Two videos were coded for each student, one at the beginning of participation in the intervention and one from the end of participation in the intervention. I hypothesized that results from t-tests used to examine change in score over time would support significant

differences in score, with variation among the seven comprehension constructs listed on the SOLR.

Results from the t-tests used to examine change in score across participants showed that the constructs on the SOLR were not sensitive to change over time. Discussion explores the lack of significant findings, including discussion of the video data used in the study, which was initially collected for purposes other than capturing comprehension behaviors. Examination of the video data and discussion supports that the SOLR likely is sensitive to change, and future research should include a plan for collecting data that eliminates variability that contributes to differences in student opportunities to respond across videos. Additionally, individual student data shows that for three students, there was improvement in all comprehension constructs over time, for four students, there was a decrease in score across most constructs; and for five students, score across construct remained relatively stable.

Conceptual Framework

The conceptual framework that guided the three papers for this dissertation is the Simple View of Reading (Gough & Tunmer, 1986). This framework describes the need for the development of both listening comprehension and word recognition processes in order to become a strong reader. Through this framework, reading comprehension is described as the product of language development and word recognition processes, and too little of one or the other results in a deficit of reading comprehension. When word recognition processes, also known as decoding ability, are weak, students experience difficulty recognizing words on a page and have fewer cognitive processes available to devote to word meaning (Adams, 1990; Perfetti, 1985; Gough & Tunmer, 1986). Additionally, when language development, also known as oral language comprehension processes, is weak, students struggle to determine the meaning of words even if they are able to decode what is on the page (Gough & Tunmer, 1986; Scarborough, 1990). In order to become a reader with developed comprehension skills, students must develop both their language development (oral language) and word recognition processes (decoding skills).

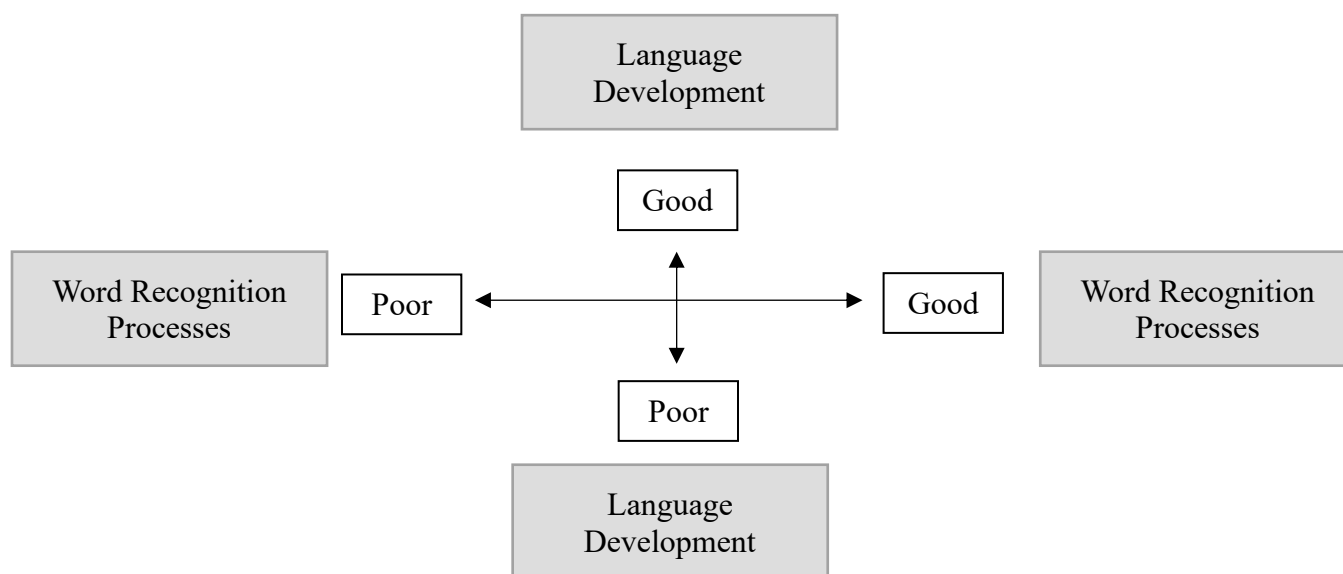


Figure 1. Simple view of reading (Gough & Tunmer, 1986)

Chapter II: A Synthesis of Reading Comprehension Measures for Students with Autism Spectrum Disorder and Low IQ

The ultimate goal of reading is to develop comprehension, as reading comprehension is a foundational skill for acquiring knowledge, engaging with community, and achieving success in the workplace (Castles et al., 2018). Researchers have historically considered reading comprehension to be one of the most important academic skills learned in school for all students (Mastropieri & Scruggs, 1997; Chiang & Lin, 2007). Bursuck and Darner (2011) describe reading comprehension as an active process that occurs when a reader gathers meaning from written text. For all students, including those with intensive needs, success in almost all academic content areas is dependent upon reading comprehension (Mims et al., 2012). Reading comprehension is also an important skill to develop for successfully navigating everyday life outside the classroom, for typically developing individuals and for individuals with disabilities, as reading comprehension is necessary for success in everyday tasks such as following steps, reading grocery lists, navigating directions and more (Mims et al., 2012).

There are several skills that must be acquired by students to be successful reading comprehenders; among these skills are behaviors such as word recognition and listening comprehension (Gough & Tunmer, 1986), along with oral language and engagement during literacy instruction (Carver, 2003; Swanson & O'Connor, 2009). Gough and Tunmer's (1986) Simple View of Reading models the manner in which word recognition and listening comprehension interact to result in reading comprehension. In other words, reading comprehension is the product of listening comprehension and word recognition, and in order for students to be successful in reading comprehension, they must have adequate skill in both domains. Too little of either word recognition or listening comprehension will result in reading

comprehension deficits. Research suggests that a majority of students who struggle with reading comprehension have a weakness in decoding ability (Nation et al., 1999). Additionally, other empirical literature suggests that for some students, weakness in comprehension may be due to deficits in areas other than decoding, such as language comprehension, background knowledge, vocabulary, and other general cognitive factors (Castles et al., 2018; Joshi & Aaron, 2000; Leach et al., 2003).

Emerging research supports that students with low IQ are able to develop reading skills in response to comprehensive instruction designed to meet the specific needs of students who demonstrate slower growth (Allor et al., 2014; Browder et al., 2008; Connor et al., 2014). Although recent research supports that students with intensive needs respond positively to reading instruction, research on reading comprehension for students with both Autism Spectrum Disorder (ASD) and low IQ is more limited, with little evidence of which interventions are effective and which measures are reliable for this population (El Zein et al., 2016; Knight et al., 2015; Turner et al., 2017). Research focusing on comprehension abilities of students with learning and other disabilities (e.g. ID, dyslexia, speech/language impairment) provides support for effective practices; however, students with ASD have not been included in many of these studies (El Zein et al., 2016; Flores & Ganz, 2007). Furthermore, reading comprehension research for students with ASD is most often limited to students with average to above average IQs. As few studies have examined reading comprehension interventions for students with ASD and low IQ, little is known about best practice for improving reading comprehension in this population (Chiang & Lin, 2007; Flores & Ganz, 2007; O'Connor & Klein, 2004; Williamson et al., 2015). Additional empirical research is necessary to determine reading practices that

effectively facilitate improvement in understanding of text for students with low IQ, particularly those with ASD (Solis et al., 2016).

Students who demonstrate deficits in reading comprehension often demonstrate deficits in listening comprehension, engagement, and other behaviors that are necessary, but not sufficient, to develop reading comprehension (Castles et al., 2018; Yuill & Oakhill, 1991). Additionally, issues with short-term memory as well as accessing prior knowledge and oral language deficiencies, many of which are often exhibited by students with ASD, are other explanations provided by researchers for deficits in reading comprehension (Hulme et al., 1997; Knott et al., 1997; Nation et al., 1999). However, without adequate measures to assess either reading comprehension or the cascading levels of comprehension behavior, such as engagement, oral language, and listening comprehension, it is difficult to distinguish the specific intervention needs of students to support their growth in reading comprehension.

Many reading assessments primarily measure word recognition, and deficits in word reading and language ability can influence individual scores on standardized assessments of reading comprehension (Nation et al., 1999; Yuill & Oakhill, 1991). Other early reading assessments primarily focus on reading speed or reading accuracy/fluency (Leach et al., 2003). Assessments used by researchers and teachers to measure reading comprehension often involve listening tasks, reading a series of sentences and making judgements about their validity, remembering a specific word (final/first) of a sentence, or reading a short passage aloud and answering inferential and recall questions (Daneman & Carpenter, 1980, 1983; Nation et al., 1999; Stothard & Hulme, 1992). Many of these described assessments either do not adequately capture the multiple components of reading comprehension, such as oral language or

engagement or are not appropriate for developing readers with limited word recognition skills, including students with disabilities (Francis et al., 2006).

Research on reading comprehension for students with ASD is limited but supports that students with ASD often have deficits in oral language, including both expressive and receptive language, which can influence comprehension abilities (Leach et al., 2003; Nation et al., 1999). Empirical research is needed to determine reading practices that effectively facilitate improvement in comprehension of text for students with ASD (El Zein et al., 2016; Flores & Ganz, 2007). Research focusing on reading practices for students with ASD has often not included students with low IQ, and there is a need for research focusing on this specific population due to their unique needs (El Zein et al., 2016; Solis et al., 2016).

Two primary theories related to the unique needs of students with ASD and their struggle with comprehension are central coherence (Turner et al., 2017) and theory of mind (ToM; Turner et al., 2017; Williamson et al., 2015). Students with strong central coherence have the ability to separate the main idea from details when reading. Students with ASD often display weak central coherence (WCC) and have difficulty making inferences and identifying the main idea when reading, which are core components of reading comprehension (Happe & Frith, 2006; Solis et al., 2016). According to Turner and colleagues (2017), individuals with WCC are overly focused on the words in text, preventing them from being able to gather the main idea. Additionally, the struggle of individuals with ASD to understand the perspectives of different characters and make inferences about their actions is explained by poor theory of mind (ToM; Turner et al., 2017; Williamson et al., 2015). ToM, necessary for strong reading comprehension, is characterized by a lack of awareness of social situations and an inability to understand intentions, actions, or feelings of someone else, all of which are common for individuals with ASD.

Weak central coherence and poor ToM are just two possible theories contributing to people with ASD struggling to develop strong reading comprehension skills. Other common characteristics of students with ASD that are likely to negatively affect reading comprehension include challenging behaviors such as resistance to novel information/instruction and self-stimulatory behaviors (Solis et al., 2016). These behaviors can be associated with difficulty teaching students with ASD and can affect their responses to assessment as well as the technical adequacy of measures used to assess growth (Ketterlin-Geller, 2008). Other common limitations associated with ASD include deficits in pragmatic language, and impaired language and communication skills (Solis et al., 2016; Turner et al., 2017). These limitations in knowledge and deficits in understanding can also affect assessment results, including formative assessments and progress monitoring data, resulting in variance in individual response (Allor et al., 2013; Jones et al., 2018; Wallace et al., 2010).

Purpose and Research Questions

This review of literature synthesizes studies that included comprehension measures used in studies of interventions including students with ASD and low IQ, to determine the most appropriate reading comprehension assessments used to track growth in this population. Through an analysis of the measures used in these studies, this paper describes the need for assessments that are sensitive enough to measure growth of reading comprehension behaviors in students with disabilities, including those with ASD and low IQ. The following section describes the search methodology that was used for the literature search. The research questions addressed in this synthesis are:

1. What measures are used to capture the comprehension growth of students with ASD and low IQ?
2. What types of comprehension constructs are captured by these measures?

Method

Inclusion Criteria

This review of literature synthesizes studies in which researchers examine comprehension interventions for students with ASD and low IQ. The reason for specifically focusing on students with low IQ is that much of the current literature focusing on reading comprehension for students with ASD focuses on students with ASD who are high functioning (formerly known as Asperger's Syndrome) and does not consider students with lower cognitive abilities. Additionally, more research on effective reading comprehension measures for students with low IQ is also necessary to plan more effective, evidence based reading interventions for this population, as accurate assessment is crucial to successful comprehension development and implementation of effective interventions (Yuill & Oakhill, 1991). A synthesis of scientific knowledge regarding reading comprehension assessments for students with ASD and low IQ would assist researchers and practitioners in planning more effective comprehension interventions for this population.

Five inclusion criteria guided this search process. First, the studies must have been published between 2004 and 2019, to stay current with the most recent mandates included in the Individuals with Disabilities Education Improvement Act (IDEIA; 2004) concerning least restrictive environment (LRE) for students receiving special education services. Mandates regarding LRE included in the 2004 reauthorization of IDEIA require that students in special education receive as much instruction in the general education setting as possible. As a result, effective instruction for students with disabilities became a larger focus of research, as the expectation is that more students with intensive needs will be included in the general education setting (IDEIA; 2004). Second, participants in the studies had to have ASD and low IQ. To

ensure I included all studies with participants who had ASD and low IQ, I used the liberal standard that if a majority of participants had a diagnosis of ASD, their IQ scores must be at least one standard deviation below the norm (i.e. 85 and below). Studies in which IQ data was not provided for students with ASD were included if adaptive behavior scores, reading comprehension scores, or other cognitive ability assessments were provided, which served as a proxy for low cognitive functioning. Again, scores had to be at least one standard deviation below the mean or below the 50th percentile on a scaled scoring system to be considered for the purposes of this review. Detailed information about the demographics of study participants, including age, IQ, and diagnosis, is provided in Table 1. Third, studies had to be experimental in design, including single case design (SCD), experimental pilot studies (pre-post), and randomized controlled trials (RCTs). Fourth, studies had to include a measure of a reading comprehension behavior, such as passage reading comprehension, listening comprehension, or engagement during literacy instruction. For example, studies were excluded if researchers only measured teacher or parent satisfaction with interventions or perception of intervention effectiveness. Finally, studies were only included if participants were school-aged students; studies including only preschool, post-secondary or adult learners were excluded.

Search Procedure

I conducted an electronic search and included the following databases that focus on educational research: *Educational Research Information Center (ERIC)*, *Academic Search Complete*, *Education Source*, and the *American Psychological Association Storage of Journals (PsychINFO)*. The following search terms were used within the databases (*reading comprehension*) or (*reading*) and (*intervention*) or (*instruction*) or (*processes*) and (*autism**) or (*ASD*) and (*assessment*) or (*measure*) or (*progress monitoring*). Articles included only empirical

studies published in peer-reviewed journals. The initial search resulted in 163 unique abstracts which were reviewed to determine which articles met the inclusion criteria. This initial screening narrowed the search from 163 to 11 studies (Bethune & Wood, 2013; El Zein et al., 2016; Flores & Ganz, 2007; Flores et al., 2013; Head et al., 2004; Kim et al., 2018; Kimhi et al., 2018; O'Connor and Klein, 2004, Turner et al., 2017; Whalon et al., 2016; Whalon et al., 2015). A PRISMA chart (see Figure 1) is provided that includes information about excluded articles. I conducted an ancestral search using the references of the eleven articles that were found through the electronic search. The following five studies were found through the ancestral search and met the criteria for this review: Dodd et al., (2011); Knight et al., (2015); Mims et al., (2012); Reynhout and Carter (2008) and Williamson et al., (2015). The inclusion of these five additional articles brought the total number of studies that met criteria for this review to 16.

Finally, I coded each article based on the research questions. First, I coded each article to determine what measures were used to capture the growth of students. Particularly, I focused on whether or not the researchers created their own measure of comprehension behavior or relied on a published measure they did not create themselves. Second, I coded the articles based on five categories that I created to determine what types of comprehension constructs were captured by the measures. The coding categories are: 1) passage reading comprehension (students are responsible for reading an entire passage of text independently); 2) supported reading passage comprehension (students are supported with audio when initially exposed to the text and then read independently); 3) sentence/phrase reading comprehension (students read a sentence or two at a time); 4) literacy engagement (researchers measure the amount of time the student is focused on the reading material); and 5) listening comprehension (student listens to text that is read

aloud by an adult). See Table 2 for more information on the comprehension behaviors and measures included in the studies.

Results

Study Design

Of the articles that met criteria for this synthesis, four different study designs were implemented by the researchers: randomized controlled trial (n=2), repeated measures (n=1), pre-post exploratory pilot study (n=1), and single case design (n=12).

Study Participants

Participants of the studies ranged in age from 4-17. The number of participants included in the studies ranged from 1 participant (El Zein et al., 2016; Reynhout & Carter, 2008; Whalon et al., 2016) to 29 participants (Turner et al., 2017). Additionally, for studies which included more than one participant, either all or a majority of participants met the inclusion criteria for IQ described in the methods section. Intervention results are only reported for the students who met inclusion criteria. Additional participant demographic information is provided in Table 1.

Summary of the Interventions and their Effectiveness

In four of the studies, researchers examined the effects of published interventions; in the other 12 studies, researchers examine the effectiveness of interventions they created themselves or with a team. Of the published interventions included in the studies, three studies included Direct Instruction (DI) reading programs (Flores & Ganz, 2007; Flores et al., 2013; Head et al., 2018), and one study included an intervention called *Book Builder*, which is a supported etext that incorporates accommodations to make books more accessible to students with disabilities (Knight et al., 2015). The interventions created by researchers varied in their design but included different accommodations appropriate to meet the unique needs of students with ASD. For

example, Dodd et al., (2011) created a perspective-taking intervention designed to assist students with ASD in their understanding of characters' thoughts and feelings. Several researchers incorporated graphics into their interventions to help students with ASD visualize what was read (Bethune & Wood, 2013; El Zein et al., 2016; Kimhi et al., 2018; Williamson et al., 2015). Other common elements imbedded into interventions were prompts and reciprocal question strategies to guide focus during reading (Kim et al., 2018; Mims et al., 2004; O'Connor & Klein, 2004; Turner et al., 2017). Many of the components of the interventions, both published and researcher created, were incorporated into the daily instruction of students with ASD and low IQ in an attempt to intensify comprehension instruction. Results from 13 of the 16 studies supported the effectiveness of the intervention for the students with ASD and low IQ, including both researcher-created interventions and published interventions. Information regarding the interventions, including the focus of the intervention and main findings, is provided in Table 3. A complete synthesis of these findings is outside the scope of this manuscript (see Conner, in progress).

RQ 1: What measures are used to capture the comprehension growth of students with ID and ASD?

In 11 of the 16 studies, researchers relied only on measures they created to assess comprehension behaviors (Bethune & Wood, 2013; Flores & Ganz, 2007; Head et al., 2018; Kim et al., 2018; Knight et al., 2015; Mims et al., 2004; O'Connor & Klein, 2004; Reynhout & Carter, 2008; Whalon et al., 2015; Whalon et al., 2016; Williamson et al., 2015). In two of the 16 studies, researchers used a combination of both published measures and researcher-created measures (El Zein et al., 2016; Flores et al., 2013). Finally, in three studies, the researchers relied on only published measures to assess comprehension behaviors (Dodd et al., 2011; Kimhi et al.,

2018; Turner et al., 2017). See Table 2 for more information about the measures included in the studies.

Researcher-Created Measures

In 11 of the 16 studies, the researcher(s) created their own assessments to measure comprehension (Bethune & Wood, 2013; Flores & Ganz; 2007; Head et al., 2018; Kim et al., 2018; Knight et al., 2015; Mims et al., 2004; O'Connor & Klein; 2004; Reynhout & Carter, 2008; Whalon et al., 2015; Whalon et al., 2016; Williamson et al., 2015). Nine of these assessments were in the form of probes or a series of questions regarding what was read, which were asked to the students during or immediately following literacy instruction (Bethune & Wood, 2013; Flores & Ganz; 2007; Head et al., 2018; Kim et al., 2018; Knight et al., 2015; Mims et al., 2004; O'Connor & Klein; 2004; Whalon et al., 2015; Whalon et al., 2016; Williamson et al., 2015). The probes or questions included questions regarding facts, identification of main idea, literal recall questions, and inferences. Most responses were scored as either correct (1) or incorrect (0); however, O'Connor and Klein (2004) scored the responses to questions on a 25-point rubric. Reynhout and Carter (2008) did not rely on questions or probes, but instead tracked the amount of time the student was focused on the book. Notably, this was the only study in which the measures did not support positive effects of the intervention (Reynhout & Carter, 2008).

Bethune and Wood (2013) measured reading comprehension in their study of the effectiveness of graphic organizers by asking eight, researcher-created, literal recall questions. The questions were all presented in the form of “wh” (who, what, where, what doing) comprehension questions about information contained in a passage that the student read. Questions were scored by the researcher as either correct or incorrect.

Flores and Ganz (2007) also measured comprehension with researcher-created probes. The probes were created to measure one of each of the three strands taught in the DI intervention *Corrective Reading Thinking Basics: Comprehension Level A* (Engelmann et al., 2002) that was focused on during their study. The three strands were statement inference, using facts, and analogies. Once students demonstrated 100% accuracy of each strand across at least three data points, beginning with statement inference, they moved to the next strand.

Head et al. (2018) also created probes to measure the effectiveness of the DI intervention *Corrective Reading Comprehension: B1* (Engelmann et al., 2008). The probes were presented in the form of questions focused on strands of skills taught in the program, including reasoning skills, information skills, vocabulary skills, sentence skills, basic comprehension skills, and writing skills. Once students achieved mastery at 90% on three consecutive probes for a skill, beginning with reasoning skills, the next skill was introduced.

Kim et al. (2018) assessed reading comprehension through 10 multiple-choice questions asked to the students. The questions included those about what happened, when it happened, where it happened, who was related to the problem, why it happened, and how the problem was resolved in the book that was read during the intervention. Task engagement was also measured (a secondary dependent variable), through thirty second momentary time sampling.

Knight et al. (2015) created probes to accompany each book in the eBook curriculum that was used in their study. The probes were three vocabulary questions, three literal comprehension questions, and one application question from the corresponding science book. The students were presented with multiple choice answer responses (one of four options) that were scored 0 for incorrect and 1 for correct. Questions were read aloud to the students through the text to speech application on the computer.

Mims et al. (2012) created questions to accompany each biography that was presented to the students in the study. The researchers asked the students 11 total comprehension questions, 8 of which were “wh” questions (who, what, where, when, why) and 3 of which were sequencing questions (What came first? Next? Last?), and presented four response options for each question. Unprompted and correct answers were tracked and graphed. Data was analyzed for each student through visual inspection for trend, level, and variability to assess whether or not a functional relation was present.

In both studies from Whalon and colleagues (2015; 2016) which examined the effectiveness of the RECALL intervention, the researchers created their own comprehension probes. These probes were embedded into the readings, similar to dialogic reading questions. The correct responses were graphed for students at baseline, during the intervention phase, and at follow up.

Williamson et al (2015) also created comprehension questions for students in response to their graphic organizer intervention. After each session, the students were asked 10 comprehension questions by the researcher. Students were allowed to look back through the book before delivering their answer. The researcher recorded the number of correct questions which were later graphed for each student and used for visual analysis as well as percentage of non-overlapping data (PND; Scruggs et al., 1987) and Conservative Dual Criterion (CDC; Fischer et al., 2003; Swoboda et al., 2010).

O'Connor and Klein (2004) created questions corresponding to each passage that were read aloud to the students after they participated in each reading condition. During each of the four conditions, the students were responsible for reading the passage out loud (control, anaphoric cueing, rereading questions, completing cloze sentences) and then verbally responding

to questions from the researcher. Responses to questions were scored on a 25-point rubric that was created by the researcher. Questions included free retelling of the story, identifying the main idea, generating a title, detecting incongruous sentences, and four “why” or “how” questions.

Reynhout and Carter (2008) did not use question probes, but measured literacy engagement in response to a researcher delivered social story intervention by tracking how often the student looked at the book during each lesson. The lessons lasted from 3.5 to 15.8 minutes with a mean length of 10 minutes, and a checklist was used to track whether or not the student looked at the book every ten seconds (10-second partial interval recording). A second dependent variable, correct responses to comprehension questions, was also measured but considered a moderator variable by the researchers. The student responded orally to questions asked by the teacher, and responses were coded as either correct or incorrect.

Combination of Published and Researcher-Created Measures

In two of the studies, researchers used a combination of their own measure and a published measure (i.e., two measures for the DV, one published and one researcher-created) to assess change in comprehension behavior. In both of these studies, measures supported positive effects of the intervention.

El Zein et al. (2016) assessed reading comprehension with two different measures. First, the researchers developed their own curriculum-based measure (CBM). Their measure consisted of five short answer questions including three facts-based questions and two inference questions. Students responded to the questions orally and researchers graphed the percentage of correct responses after each session. The second measure, the published measure, was an oral retelling measure that followed the *Dynamic Indicators of Basic Early Literacy Skills Oral Reading Fluency* measure (DIBELS; Good & Kaminski, 2002). During this measure, students are asked

to retell a passage in one minute with as much detail as they can provide. The instructor counts the number of relevant words that are spoken by the student and that number is then recorded.

To measure comprehension growth of the participants in their study, Flores et al. (2013) used a combination of researcher-created assessments with the CBM that was included in the DI programs of focus. The first group of students participated in the DI program *Language for Learning* (LL; Engelmann & Osborne, 1999). The initial placement test for this group was a performance measure created by the researchers. The second performance measure for this group was administered to the students two weeks after instruction began and consisted of the curriculum-based assessments that were published in LL as mastery tests to be delivered after every tenth lesson. The third performance measure, a combination of several published mastery assessments from the LL curriculum, was administered two weeks after the second performance measure. The second group of students participated in the DI program *Corrective Reading Comprehension: A Thinking Basics* (CR; Engelmann et al., 2002). The first measure was the placement test created by the researchers. The second performance measure was also created by the researcher, using a similar format to the published mastery tests included in the CR program, and was delivered after two weeks of instruction. After two more weeks of instruction, the researchers administered the third performance measure, which consisted of the published CBM included in the CR program, without any modifications.

Published Measures

In three studies, researchers used published measures without any modifications as the measure of comprehension behavior (Dodd et al., 2011; Kimhi et al., 2018; Turner et al., 2017). In each of these studies, measures support a positive effect of the intervention.

Dodd and colleagues (2011) used the perspective taking score (PTS; Garcia-Perez et al., 2008) to measure oral story retell from different perspectives of different characters, which served as the dependent variable in their study of a narrative based intervention. After participants listened to the story twice, they retold the story from the perspective of the main character and then from another character. The oral retells were transcribed and evaluated by the researcher. Each retell received a score using the PTS.

In their pilot study, Kimhi and colleagues (2018) used a pre-post design to examine the effect of a modified reading intervention on the reading comprehension growth of kindergarten students with ASD. A standardized measure of reading comprehension, the Katzenberger Hebrew Language Assessment (KHLA; Katzenberger, 2009), was used to assess reading comprehension at pretest and posttest. The researchers used a Wilcoxon signed ranks test to determine the significance between the pretest and posttest scores.

Turner, Remington, and Hill (2017) used the *York Assessment of Reading for Comprehension* (YARC; Snowling et al., 2009) to assess reading comprehension for students participating in their study aiming to assess the effectiveness of a reciprocal questioning approach delivered during reading instruction. The YARC incorporates two measures of reading comprehension: 1) responses to orally presented questions of reading comprehension and 2) a summarization of the text, which is completed after the comprehension questions are answered. Students independently read passages of text that are included in the assessment before answering the questions and completing the oral summarization.

RQ 2: What types of comprehension constructs are captured by these measures?

I created five coding categories based on the measures included in the studies to evaluate comprehension constructs. Again, the five coding categories that I created to evaluate the

comprehension constructs are: 1) passage reading comprehension; 2) supported reading passage comprehension; 3) sentence/phrase reading comprehension; 4) literacy engagement; and 5) listening comprehension. Information on type of measure used and the comprehension behavior assessed is provided in Table 2.

Passage Reading Comprehension

In five studies, the comprehension construct the researchers measured was passage reading comprehension (Bethune & Wood, 2013; El Zein et al., 2006; Flores et al., 2013; O'Connor & Klein, 2004; Turner et al., 2017). In these studies, comprehension was assessed through student responses to passages that were read independently. Students were responsible for reading entire passages or chapters of text before completing the comprehension measure. The comprehension construct measured in these studies was coded as passage reading comprehension due to the requirement that the student complete the portion of book or passage reading that is later assessed by the researcher. The key characteristic of this comprehension construct was the lack of teacher/researcher support during reading.

Supported Passage Reading Comprehension

In one study, the comprehension construct the researchers measured was supported passage reading (Williamson et al., 2015). This study was coded as supported reading comprehension because the students had the option to listen to a chapter of text the first time it was presented to them while following along in their books. In this study, the students originally listened to an audio recording of the story while following along with their own version of the book, but during assessment, the students were responsible for reading independently.

Phrase or Sentence Comprehension

In four studies, the comprehension construct the researchers measured was phrase or sentence comprehension (Flores & Ganz, 2007; Head et al., 2018; Kim et al., 2018; Kimhi et al., 2018). In these studies, students were required to read short phrases or sentences independent of their teacher (or the researcher); however, the required reading was shorter than readings in interventions in which researchers measured passage reading comprehension.

Listening Comprehension

In five studies, researchers measured listening comprehension (Dodd et al., 2011; Knight et al., 2015; Mims et al., 2004; Whalon et al., 2015; Whalon et al., 2016). These studies are coded as listening comprehension because the researcher (or a computer) read aloud to the students, and students responded to probes or questions after the text was read. The students were not responsible for independent reading; rather, they listened to information that was read aloud by a researcher or teacher.

Literacy Engagement and Listening Comprehension

In one study, the primary comprehension construct the researchers measured was literacy engagement (Reynhout & Carter, 2008). In this study, the researchers also included a focus on a secondary comprehension behavior, listening comprehension. The researchers measured literacy engagement by tracking the amount of time that the student spent looking at the book during instruction, which was the first dependent variable measured by the researchers. The researchers also tracked responses listening comprehension through questions that were asked to the student after the researcher read the text aloud, which were reported in the study as the secondary dependent variable.

Discussion

The purpose of this synthesis was to examine current research on reading comprehension assessments used in interventions for students with ASD and low IQ. The ultimate goal of reading, for all students, is to develop comprehension, as reading comprehension is necessary to participate successfully in academic content areas and to navigate life outside of school, such as home, the community, or the work place (Mastropieri & Scruggs, 1997; Chiang & Lin, 2007; Castles et al., 2018). Again, the development of a comprehension measure that is inclusive of students with disabilities is crucial to reading development for this population as poor assessments of reading comprehension can potentially discourage willingness to read and cause teachers to implement poor practices (Nation et al., 1999; Yuill & Oakhill, 1991).

The research questions for this synthesis focus on the measures used to assess comprehension in the intervention studies and the comprehension constructs that those measures assess. The first research question examined the types of measures used to capture comprehension growth that were included in the studies. The measures included both measures created by the researchers and published measures. Many of the measures were in the form of fact or retell probes, and some included questions regarding inferences, understanding of character point of view, or oral retell of the main point. The second research question examined the comprehension constructs that were captured in the measures. Of the 16 articles in this review, comprehension behaviors included passage reading comprehension, supported passage reading comprehension, phrase/sentence comprehension, literacy engagement, and listening comprehension. The findings, including the measures and constructs, are discussed in further detail below, along with limitations of this study and directions for future research.

RQ 1: What measures are used to capture the comprehension growth of students with ID and ASD?

With respect to the measures researchers used for the comprehension behavior in their studies, 11 of the studies included in this review contained measures that were created by the researchers for their studies, two studies included a mix of published and researcher created measures, and three studies contained only published measures (see Table 2).

Two of the studies used some form of oral retell as a measure of comprehension behavior (El Zein et al., 2016; Turner et al., 2017). Oral retell was scored based on how accurately the student, in their own words, could describe what was read, timing for responses was limited to one minute (El Zein et al., 2016) and was untimed in another (Turner et al., 2017). Oral retelling is an established method for assessing comprehension; however, procedures for oral retell measures tend to vary among tools (Reed & Vaughn, 2011). Additionally, in their synthesis examining protocols for oral retell measures, Reed and Vaughn (2011) found that oral retell was a more accurate measure of reading comprehension in elementary-aged students than in older students. Though the researchers in these studies did not use standardized assessments, they described their procedures for retelling with enough detail that they could be replicated by other researchers.

Other measures used by researchers included CBM that were included in DI programs and unpublished assessments/probes created by the researcher, both of which include questions regarding facts and inferences from reading. As many current early reading assessments focus on decoding/reading speed/fluency (Leach et al., 2003), many of the researchers turned to their own measure of comprehension behavior to track the dependent variable in their study. While no two studies contain the same measure for assessing comprehension behavior, researchers who created their own unpublished measure often created a series of questions or probes specific to the passage or phrase that was read. Researchers created those probes to specifically align to the

intervention and serve as proximal measures. While the measures used may not be transferable to other reading interventions, a strong rationale for this approach is that these measures would likely be more sensitive to change than standardized measures. However, there is a need for a measure of comprehension behavior that can be used across different interventions to assess growth, so that results can be synthesized across studies.

RQ 2: What types of comprehension constructs are captured by these measures?

Within the 16 studies located for this review, five different comprehension behaviors were assessed through the measures used in the interventions: passage reading comprehension, supported passage reading comprehension, phrase/sentence comprehension, literacy engagement, and listening comprehension. In five of the studies, researchers measured passage reading comprehension, and in one study, the researcher measured supported passage reading comprehension. In the other studies, researchers measured varying comprehension behaviors, such as phrase/sentence comprehension (N=4), listening comprehension (N=5), and literacy engagement (N=1).

In four studies, researchers measured phrase or sentence comprehension, which is comparable to passage reading comprehension; however, the students are required to read less independent material before their comprehension is assessed than during passage reading comprehension. In these interventions, the students read either a short phrase or one sentence at a time and were then asked questions regarding the information in that sentence or short phrase. Although the results of studies that measure phrase or sentence comprehension and those that measure passage comprehension are not easily comparable, one could consider sentence/phrase comprehension as a skill that must be mastered by students before they can be expected to comprehend an entire passage successfully. The ability to successfully comprehend short phrases

is possibly a building block to successful comprehension of longer passages. If a student is unable to comprehend an entire passage, the next step would be to assess whether or not the student is able to comprehend short phrases.

In addition to passage reading comprehension and sentence/phrase comprehension, researchers also measured literacy engagement in one study and listening comprehension in five studies. The theoretical frame the Simple View of Reading (Gough & Tunmer, 1986) describes listening comprehension, or language comprehension, as an essential component of strong reading comprehension. In this model, a reader must develop both language comprehension and word reading to excel in comprehension. Additionally, literacy engagement is also a necessary skill to acquire for development of strong reading comprehension, as students must pay attention to the information they process while reading in order to comprehend (Carver, 2003; Swanson & O'Connor, 2009). Similar to sentence/phrase comprehension, listening comprehension and literacy engagement can also be considered building blocks to developing reading comprehension, and many students who struggle with reading comprehension often have deficits in constructs such as listening comprehension or engagement (Castles et al., 2018; Yuill & Oakhill, 1991). While the outcome measures used for listening comprehension, engagement, and reading comprehension are too different to compare on the same scale, they are related. If reading comprehension is weak, teachers or researchers can assess word reading and language/listening comprehension separately to find the area the student needs the most intervention, as these skills are necessary for the development of reading comprehension (Carver, 2003; Swanson & O'Connor, 2009; Yuill & Oakhill, 1991).

Within the sixteen studies, five different comprehension behaviors were identified. While these comprehension behaviors are distinct, they are all related and must be developed in order

for students to acquire passage-reading comprehension. The behaviors measured by the researchers are cascading skills that contribute to the development of successful reading comprehension. The need for a measure of comprehension that includes all of the behaviors identified (i.e. engagement, listening comprehension, sentence comprehension, etc.) is clear. A measurement tool that includes multiple comprehension behaviors that is inclusive of students with low IQ would allow researchers to compare outcomes of multiple interventions to determine which are most effective.

Limitations

Many of the measures used by researchers whose studies were included in this synthesis were created to specifically align to the intervention and adequately capture change in comprehension behavior. While direct alignment improves sensitivity to change, it is also a potential limitation, as growth may not transfer to improvements in reading comprehension in other contexts. However, researchers and teachers can continue to create probes and questions specific to comprehension behavior or intervention as a method to assess the success of potential interventions until a more systematic measure is created.

Additional limitations related to the search include limiting articles to those published after 2004, due to mandates with LRE. It is possible that by extending the search to include studies published before 2004 additional articles would have been located that met search criteria. In addition, the exclusion of studies that focused on students with “high functioning” ASD is recognized as a potential limitation and expanding the search to include those types of participants could potentially result in more studies focusing on comprehension behaviors and a wider range of measures and behaviors.

A final limitation is that I created my own coding scheme for describing and categorizing the measures used to assess comprehension in the studies. I addressed this limitation in the methods section with a clear rationale for the coding scheme and a detailed description of the coding process.

Implications for Future Research

In designing comprehension studies with students with ASD and low IQ, researchers should consider the broad range of skills that relate to reading comprehension, including phrase/sentence comprehension, listening comprehension, and literacy engagement. Measures should be considered that address this developmental sequence. These comprehension behaviors are connected, meaning they build upon one another and are all integral components of strong reading comprehension. Researchers should consider the developmental order of these skills when selecting and designing measures and future research may further inform our understanding of comprehension development, particularly for students with intensive needs. Future research should focus on a systematic approach to comprehension assessment that can be individualized based on student need, which is often necessary when working with students with ASD and low IQ.

The interventions included in the study were designed to target the needs of students with ASD who often struggle to make connections, identify the main idea, and understand characters (Turner et al., 2017; Williamson et al., 2015). Interventions that have been demonstrated to be successful with other students (e.g., students with LD) need to be examined with students with ASD and low IQ. Researchers should consider how these interventions might need to be adapted and potential new interventions for these students. In order to accurately assess intervention

effectiveness, reliable and valid reading comprehension measures that are sensitive enough to capture growth are necessary.

References

- Allor, J. H., Mathes, P. G., Roberts, J. K., Cheatham, J. P., & Otaiba, S. A. (2014). Is scientifically based reading instruction effective for students with below-average IQs? *Exceptional Children*, 80, 287-306. doi: 10.1177/0014402914522208
- *Bethune, K. S., & Wood, C. L. (2013). Effects of wh-question graphic organizers on reading comprehension skills of students with autism spectrum disorders. *Education and Training in Autism and Developmental Disabilities*, 48, 236-244.
- Brown, L., & Leigh, J. E. (1986). Adaptive Behavior Inventory. Austin, TX: PRO-ED.
- Browder, D. M., Ahlgrim-Dezell, L., Courtade, G., Gibbs, S. L., & Flowers, C. (2008). Evaluation of the effectiveness of an early literacy program for students with significant developmental disabilities. *Exceptional Children*, 75, 33-52.
doi:10.1177/001440290807500102
- Bursuck, W. D., & Darner, M. (2011). *Reading instruction for students who are at risk or have disabilities* (2nd ed.). Boston, MA: Pearson Education, Inc.
- Carver, R. P. (2003). The highly lawful relationships among pseudoword decoding, word identification, spelling, listening, and reading. *Scientific Studies of Reading*, 7, 127-154.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19, 5-51.
[doi:10.1177/1529100618772271](https://doi.org/10.1177/1529100618772271)
- CAST (2014). Transforming education through universal design for learning. Retrieved from <http://www.cast.org/>

- Chiang, H., & Lin, Y. (2007). Reading comprehension instruction for students with autism spectrum disorders: A review of the literature. *Focus on Autism and Other Developmental Disabilities, 22*, 259-267.
- Conner, C. (in progress). The effectiveness of reading comprehension interventions designed for students with autism spectrum disorder and low IQ.
- Connor, C. M., Alberto, P. A., Compton, D. L., & O'Connor, R. E. (2014). Improving Reading Outcomes for Students with or at Risk for Reading Disabilities: A Synthesis of the Contributions from the Institute of Education Sciences Research Centers. NCSER 2014-3000. *National Center for Special Education Research*.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Memory and Language, 19*, 450.
- Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(4), 561.
- *Dodd, J. L., Ocampo, A., & Kennedy, K. S. (2011). Perspective taking through narratives: An intervention for students with ASD. *Communication Disorders Quarterly, 33*, 23-33.
doi:10.1177/1525740110395014
- Dunn, L. M., & Dunn, L. M. (1997). Peabody Picture Vocabulary Test. Circle Pines, MN: American Guidance Service.
- *El Zein, F., Solis, M., Lang, R., & Kim, M. K. (2016). Embedding perseverative interest of a child with autism in text may result in improved reading comprehension: A pilot study. *Developmental Neurorehabilitation, 19*, 141-145.
doi:10.3109/17518423.2014.915893

Engelmann, S., & Osborn, J. (1999). *Language for Learning*. Columbus, OH: SRA/McGraw-Hill.

Engelmann, S., Osborne, S., & Hanner, S. (2008). *Corrective Reading Comprehension: B1*. Columbus: SRA/McGraw-Hill.

Engelmann, S., Haddox, P., Hanner, S., & Osborn, J. (2002). *Corrective reading thinking basics: Comprehension level A*. Columbus, OH: SRA/McGraw-Hill.

*Flores, M. M., & Ganz, J. B. (2007). Effectiveness of direct instruction for teaching statement inference, use of facts, and analogies to students with developmental disabilities and reading delays. *Focus on Autism and Other Developmental Disabilities*, 22, 244-251.
doi:10.1177/10883576070220040601

*Flores, M. M., Nelson, C., Hinton, V., Franklin, T. M., Strozier, S. D., Terry, L., & Franklin, S. (2013). Teaching reading comprehension and language skills to students with autism spectrum disorders and developmental disabilities using direct instruction. *Education and Training in Autism and Developmental Disabilities*, 48, 41-48.

Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading*, 10, 301–322.

Garcia-Perez, R. M., Hobson, R. P., & Lee, A. (2008). Narrative role-taking in autism. *Journal of Autism and Developmental Disorders*, 38, 156–168.

GL Assessment (2011). *The British Abilities Scales 3rd Edition (BAS-III)*. London: GL Assessment

Good, R. H., Kaminski, R. A. (2002) DIBELS oral reading fluency passages for first through third grades (Technical Report 10). Eugene, OR: University of Oregon.

Gough, P. B., & Tumner, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6-10.

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 36, 5-25.
doi:10.1007/s10803-005-0039-0

*Head, C. N., Flores, M. M., & Shippen, M. E. (2018). Effects of direct instruction on reading comprehension for individuals with autism or developmental disabilities. *Education and Training in Autism and Developmental Disabilities*, 53, 176-191.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a reintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23, 1217–1232.

Individuals With Disabilities Education Improvement Act of 2004, 20 U.S.C. 1400 et seq

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, 21, 85-97.

Jones, F. G., Gifford, D., Yovanoff, P., Al Otaiba, S., Levy, D., & Allor, J. (2019). Alternate assessment formats for progress monitoring students with Intellectual Disabilities and below average IQ: An exploratory study. *Focus on Autism and Other Developmental Disabilities*, 34, 41-51.

Katzenberger, I. (2009) *The Katzenberger Language Assessment for Preschool Children*.

Mishmar Hasharon: Guy Agencies [In Hebrew].

- Ketterlin-Geller, L. R. (2008). Testing students with special needs: A model for understanding the interaction between assessment and student characteristics in a universally designed environment. *Educational Measurement: Issues and Practices*, 27, 3–16.
- *Kim, S. Y., Rispoli, M., Lory, C., Gregori, E., & Brodhead, M. T. (2018). The effects of a shared reading intervention on narrative story comprehension and task engagement of students with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 48, 3608-3622. doi:10.1007/s10803-018-3633-7
- *Kimhi, Y., Achtarzad, M., & Tubul-Lavy, G. (2018). Emergent literacy skills for five kindergartners with autism spectrum disorder: a pilot study. *Journal of Research in Special Educational Needs*, 18, 211-221. doi: 10.1111/1471-3802.12406
- *Knight, V. F., Wood, C. L., Spooner, F., Browder, D. M., & O'Brien, C. P. (2015). An exploratory study using science eTexts with students with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities*, 30, 86-99.
[doi:10.1177/1088357614559214](https://doi.org/10.1177/1088357614559214)
- Knott, R., Patterson, K., & Hodges, J. R. (1997). Lexical and semantic binding effects in short-term memory: Evidence from semantic dementia. *Cognitive Neuropsychology*, 14, 1165–1216.
- Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology*, 95, 211. doi:10.1037/0022-0663.95.2.211
- Mastropieri, M. A., & Scruggs, T. E. (1997). Best practices in promoting reading comprehension in students with learning disabilities 1976 to 1996. *Remedial and Special Education*, 18, 198-213. [doi:10.1177/074193259701800402](https://doi.org/10.1177/074193259701800402)

- *Mims, P. J., Hudson, M. E., & Browder, D. M. (2012). Using read-alouds of grade-level biographies and systematic prompting to promote comprehension for students with moderate and severe developmental disabilities. *Focus on Autism and Other Developmental Disabilities*, 27, 67-80. [doi:10.1177/1088357612446859](https://doi.org/10.1177/1088357612446859)
- Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Psychology*, 73, 139-158.
- *O'Connor, I. M., & Klein, P. D. (2004). Exploration of strategies for facilitating the reading comprehension of high-functioning students with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 34, 115-127.
[doi:10.1023/B:JADD.0000022603.44077.6b](https://doi.org/10.1023/B:JADD.0000022603.44077.6b)
- Reed, D. K., & Vaughn, S. (2012). Retell as an indicator of reading comprehension. *Scientific Studies of Reading*, 16, 187-217. [doi:10.1080/10888438.2010.538780](https://doi.org/10.1080/10888438.2010.538780)
- *Reynhout, G., & Carter, M. (2008). A pilot study to determine the efficacy of a social story intervention for a child with autistic disorder, intellectual disability and limited language skills. *Australasian Journal of Special Education*, 32, 161-175.
[doi:10.1017/S1030011200025823](https://doi.org/10.1017/S1030011200025823)
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single subject research: Methodology and validation. *Remedial and Special Education*, 8, 24–33.
- Semel, E., Wiig, E., & Secord, W. (2003). Clinical Evaluation of Language Fundamentals—4th Edition. San Antonio, TX: Pearson.

- Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. *Reading and Writing, 4*, 245–256.
- Snowling, M.J., Stothard S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K. & Hulme, C. (2009). York Assessment of Reading for Comprehension – Passage Reading. London: GL Assessment
- Solis, M., El Zein, F., Vaughn, S., McCulley, L. V., & Falcomata, T. S. (2016). Reading comprehension interventions for students with autism spectrum disorders: An alternating treatments comparison. *Focus on Autism and Other Developmental Disabilities, 31*, 284-299. [doi:10.1177/1088357615583464](https://doi.org/10.1177/1088357615583464)
- Swanson, H. L., & O'Connor, R. (2009). The role of working memory and fluency practice on the reading comprehension of students who are dysfluent readers. *Journal of Learning Disabilities, 42*, 548-575. doi:[10.1177/0022219409338742](https://doi.org/10.1177/0022219409338742)
- Swoboda, C. M., Kratochwill, T. R., & Levin, J. R. (2010). Conservative Dual-Criterion Method for single-case research: A guide for visual analysis of AB, ABAB, and multiple-baseline designs (Working Paper No 2010- 13). Madison: Wisconsin Center for Education Research. Retrieved from http://www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2010_13.php
- *Turner, H., Remington, A., & Hill, V. (2017). Developing an intervention to improve reading comprehension for children and young people with autism spectrum disorders. *Educational and Child Psychology, 34*, 13-26.
- Wallace, T., Tichá, R., & Gustafson, K. (2010). Technical characteristics of general outcome measures (GOMs) in reading for students with significant cognitive disabilities. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 26*, 333–360.

- *Whalon, K., Hanline, M. F., & Davis, J. (2016). Parent implementation of RECALL: A systematic case study. *Education and Training in Autism and Developmental Disabilities*, 211-220.
- *Whalon, K., Martinez, J. R., Shannon, D., Butcher, C., & Hanline, M. F. (2015). The impact of reading to engage children with autism in language and learning (RECALL). *Topics in Early Childhood Special Education*, 35, 102-115. doi: 10.1177/0271121414565515
- *Williamson, P., Carnahan, C. R., Birri, N., & Swoboda, C. (2015). Improving comprehension of narrative using character event maps for high school students with autism spectrum disorder. *The Journal of Special Education*, 49, 28-38. [doi:10.1177/0022466914521301](https://doi.org/10.1177/0022466914521301)
- Yuill, N., & Oakhill, J. V. (1991). *Children's problems in text comprehension*. Cambridge, UK: Cambridge Univ. Press.
- Zimmerman, I., Steiner, V., & Pond, R. (2011). *Preschool Language Scales—Fifth Edition (PLS-5)*. San Antonio, TX: The Psychological Corporation.

Table 1*Participant Demographic Information*

Study	N	Age	Diagnosis	IQ
Bethune & Wood (2013)	3	8, 10, 10	All ASD	94, 67, 90
Dodd et al. (2011)	18	3-5th grade (9 - 12 years.); 11 boys, 7 girls	All ASD	No IQ info
El Zein et al. (2016)	1	8 boy	ASD	No IQ info
Flores et al. (2013)	18	1-7th grade; all boys; group 1) 8-13 years; group 2) 7-9 years	group 1: 7 with ASD, 4 with ASD and ID or OHI; group 2: 3 with DD and 4 are ASD	14 students have IQ below 85; 4 students are 86 and above
Flores & Ganz (2007)	4	14 boy, 11 girl, 13 girl, 10 girl	ASD, ASD, ID, ADHD	No IQ, no IQ, 57, 75
Head et al. (2018)	3	10 boy, 16 girl, 14 boy	ASD, ID, ASD	82, 64, 62
Kim et al. (2018)	3	6, 7, and 8 year old boys	ASD	No IQ, 2-7 hours per day in ASD clinic receiving behavior therapy
Kimhi et al. (2018)	5	5-7 years old	ASD	Use verbal-mental age; 90, 63, 64, 95, 63
Knight et al. (2015)	4	11 boy, 11 girl, 12 boy, 14 girl	All ASD	55, 53, 63, 67
Mims et al. (2004)	4	14 girl, 14 boy, 13 boy, 12 boy	All ASD and ID	42, unable to test or score others
O'Connor & Klein (2004)	20	mean age was 15.11, SD .99; 19 boys 1 girl	10 with ASD, 6 Asperger's, 4 PDDNOS,	Mean IQ is 88; 14 students between 72-80
Reynhout & Carter (2008)	1	8 girl	ASD	No IQ
Turner et al. (2017)	29	Avg. age 13 years and 6 months; 8 boys and 7 girls in intervention; 12 boys and 2 girls in control	ASD (some have Asperger's)	No IQ
Whalon et al. (2015)	4	4, 4, 4, 5 years old; all male	All have ASD and DD	No IQ, below 2th percentile on PLS-5 measure
Whalon et al. (2016)	1	4 year old boy	ASD	64 total language score on PLS-5 (greater than 1 SD below the mean)
Williamson et al. (2015)	3	17, 16, 16 all boys	All ASD	No IQ, 76, no IQ

Table 2*Comprehension Behavior and Measure Information*

Study	Study Type	Comprehension Behavior	Measure	Researcher-created (Y/N)	Mastery or Analysis	Who is reading?
Bethune & Wood (2013)	SCD	Passage reading comprehension	Responses to 8 recall questions	Yes	7 out of 8 questions on 3 sessions	Student
Dodd et al. (2011)	RCT	Listening comprehension	Oral retell of story	No; PTS (Garcia-Perez et al., 2008)	Cohen's <i>d</i> , differences between means at pre and post test	Researcher
El Zein et al. (2016)	SCD	Passage reading comprehension	5 responses to reading comp questions (3 facts; 2 inferences); number of WMP in oral story retell	Yes and for reading comprehension; No for oral retell, (DIBELS; Good & Kaminski, 2002)	Comparing mean scores during PI and non-PI conditions; WMP on oral retell	Student
Flores et al. (2013)	Repeated Measures	Passage reading comprehension	Reading assessments presented as mastery tests	Yes and no – researcher-created CBM & CBM from LL and CR programs	ANOVA	Students
Flores & Ganz (2007)	SCD	Sentence/phrase reading comprehension	Probes focused on inferences, using facts, & analogies	Yes	100% on three consecutive probes	First researcher, then student
Head et al. (2018)	SCD	Sentence/phrase reading comprehension	Probes covering parts of speech and comparison of sentences	Yes	90% or better on 3 consecutive probes	Students

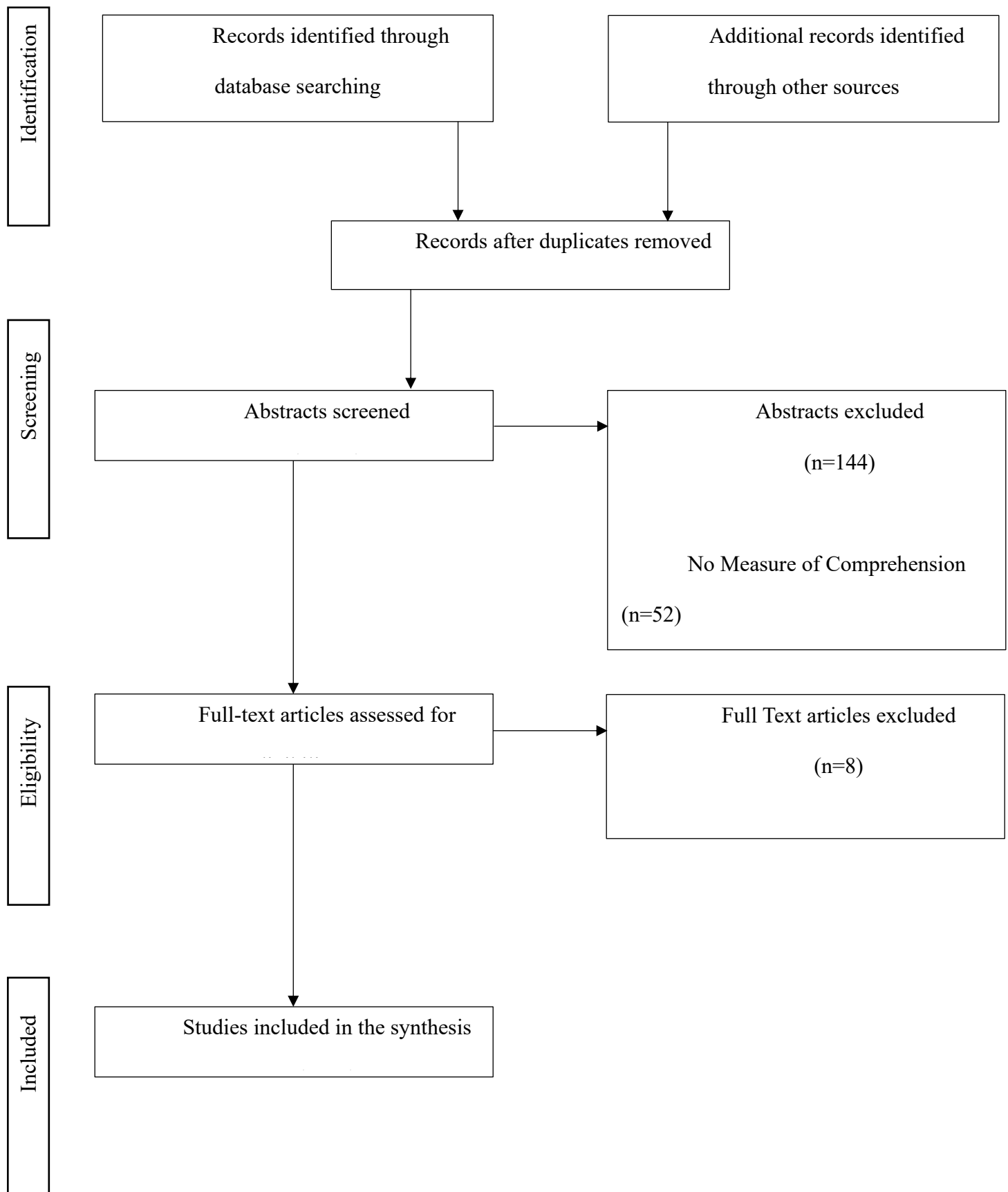
Study	Study Type	Comprehension Behavior	Measure	Researcher-created (Y/N)	Mastery or Analysis	Who is reading?
Kim et al. (2018)	SCD	Passage reading comprehension and literacy engagement	10 multiple choice questions (comp.); 30 second time sampling (engagement)	Yes	Correct, unprompted responses are recorded	The student
Kimhi et al. (2018)	Pre-Post pilot study	Passage comprehension and oral language	KHLA and oral story retelling	No	Difference scores with nonparametric Wilcoxon signed ranks test	First teacher, then choral w/ student
Knight et al. (2015)	SCD	Listening comprehension	Probes: 3 vocabulary, 3 literal comprehension, 1 application	Yes	Correct responses are graphed for each session	Computer reads text aloud
Mims et al. (2004)	SCD	Listening comprehension	Answering 8 "wh" questions (who, what, where, when, why) and 3 sequence questions	Yes	Correct responses were graphed (165 across 5 books)	Researcher
O'Connor & Klein (2004)	SCD	Passage reading comprehension	Responses to questions: retelling, main idea, title, inferences	Yes	Responses scored on a 25pt rubric created by the researchers	Students
Reynhout & Carter (2008)	SCD	Literacy engagement	1) Looking at the book 2) responses to comprehension questions	Yes	PND calculated using the data points from the three phases	Researcher
Turner et al. (2017)	RCT	Passage reading comprehension	YARC (Snowling et al., 2009)	No; YARC (Snowling et al., 2009)	ANOVA	Students
Whalon et al. (2015)	SCD	Listening comprehension	Correct responses, incorrect responses, verbal initiations, and	Yes	Correct responses were graphed, <i>Tau-U</i> also used	Researcher

Study	Study Type	Comprehension Behavior	Measure	Researcher-created (Y/N)	Mastery or Analysis	Who is reading?
			nonverbal initiations were all graphed			
Whalon et al. (2016)	SCD	Listening comprehension	Unprompted correct responses to comprehension questions	Yes	Correct responses to comprehension questions are graphed, NAP=1.0	Researcher
Williamson et al. (2015)	SCD	Supported passage reading comprehension	10 questions per strand	Yes	Correct responses are graphed for each session	First reading is supported by audio

Table 3*Intervention Information*

Study	Researcher-created (Y/N)	Focus or name of intervention	Findings
Bethune & Wood (2013)	Yes	Graphic organizer focused on “Who, Where, What, and What doing”	Functional relation between reading comprehension and graphic organizer intervention for all three students
Dodd et al. (2011)	Yes	Perspective taking intervention (focuses on state of mind and emotions of characters) and NBLI intervention (business as usual, focused on story elements)	Intervention group outperforms control; Cohen’s <i>d</i> at posttest was .96 for the PTI group and .41 for NBLI group
El Zein et al. (2016)	Yes	Altered stories involving the pervasive interest of the student (cars)	Mean score for PI condition: 70%; mean score for control condition: 38%; retell for PI (M = 8 WPM); retell for control (M = 4 WPM)
Flores et al. (2013)	No	DI intervention: <i>Language for Learning</i> (LL) & <i>Correcting Reading Comprehension: A thinking basics</i> (CR)	ANOVA suggests significant improvement over time; CR = Wilk's A = 0.075, $F(2, 9) = 55.37$, $p < .01$, multivariate $\eta^2 = .93$; LL = Wilk's A = .014, $F(1, 5) = 173.1$, $p < .01$, multivariate $\eta^2 = .99$.
Flores & Ganz (2007)	No	DI Intervention: <i>Corrective Reading Thinking Basics: Comprehension Plus</i>	Functional relation between comprehension and intervention phase was found for all students
Head et al. (2018)	No	DI intervention: <i>Corrective Reading Comprehension: B1</i>	Functional relation between comprehension and DI intervention was found for each participant
Kim et al. (2018)	Yes	Modified narrative text including topic anticipation, dynamic reading, and story retelling	Functional relation between comprehension, engagement and the researcher-created intervention

Study	Researcher-created (Y/N)	Focus or name of intervention	Findings
Kimhi et al. (2018)	Yes	Modified a preexisting intervention to include visual aids, language supports, and cognitive modifications	Significant gains in meaning related skills at post-test
Knight et al. (2015)	No	Supported eText: <i>Book Builder</i>	Functional relation between comprehension outcomes and the e-text was found
Mims et al. (2004)	Yes	System of least intrusive prompts embedded in modified text	Functional relation between listening comprehension and the intervention; three students were able to generalize these skills to new stories
O'Connor & Klein (2004)	Yes	Three created conditions cloze sentences, anaphoric cuing, reciprocal questioning (involving pre-reading questions) and one business as usual condition	Anaphoric cuing significantly increased students' passage comprehension, $F(1,19) = 5.60$, $p = .03$ $\eta^2 = .42$.
Reynhout & Carter (2008)	Yes	Individualized social story intervention	No functional relation between participation and engagement or comprehension
Turner et al. (2017)	Yes	Reciprocal questioning approach	YARC increased significantly in the intervention over the control condition $F(2,26)=12.53$; $p<.001$; effect size of $r=0.49$
Whalon et al. (2015)	Yes	Reading to Engage Children with Autism in Language and Learning (RECALL) is a shared reading routine created by the researchers that incorporates strategies and prompts	Functional relation between correct responses, spontaneous responses and RECALL for all participants, functional relation between initiations and RECALL for 3 of the 4 participants
Whalon et al (2016)	Yes	Reading to Engage Children with Autism in Language and Learning (RECALL)	Functional relation between correct, unprompted responses to comprehension questions and participation in the RECALL intervention
Williamson et al. (2015)	Yes	Character event map	Functional relation between comprehension growth and intervention for all students; PND for all students was 100%

Figure 1. *Screening Procedures*

Chapter III: Examining the Technical Adequacy of the Systematic Observation of Language and Reading

Many early reading assessments primarily measure aspects of reading such as word recognition, reading speed, or reading fluency and do not measure the reader's understanding of text (Leach et al., 2003; Yuill & Oakhill, 1991). Assessments of reading comprehension often involve reading a series of sentences and making judgements about their validity or reading a short passage aloud and answering inferential and recall questions through multiple choice or fill in the blank responses (Daneman & Carpenter, 1980, 1983; Fletcher, 2006; Nation et al., 1999; Stothard & Hulme, 1992). Many of these types of assessments either do not fully capture reading comprehension or are not appropriate for beginning readers, including students with disabilities, as these approaches to the measurement of reading comprehension often have narrow response formats (Fletcher, 2006; Francis, 2006). While the assessment of comprehension can be difficult as any single attempt to measure comprehension is often one-dimensional, results from assessments are used to make inferences about how well a student, including a student with disabilities, comprehends written material (Cutting & Scarborough, 2006; Fletcher, 2006; Francis et al., 2006). The purpose of this paper is to introduce a new measure for assessing reading comprehension behaviors that is inclusive of students with disabilities and to examine the technical adequacy of this measure.

Children with both Intellectual Disability (ID) and Autism Spectrum Disorder (ASD) have historically had limited opportunities surrounding reading instruction (Katims, 2001); however, these students are now being held to higher standards due to federal policies. These policies, such as No Child Left Behind (NCLB, 2002) and the Every Student Succeeds Act (ESSA, 2015), mandate the tracking of yearly adequate progress for all students and raising the

expectations for literacy achievement for students with disabilities (Browder & Spooner, 2014; Katims, 2001). Encouragingly, emerging research supports the ability of students with ID to learn to read in response to comprehensive instruction designed to meet the specific needs of students who demonstrate slower growth (Allor et al., 2014; Browder, Ahlgrim-Dezell et al., 2008; Connor et al., 2014).

Although recent research supports that students who require intensive supports respond positively to reading instruction and are able to read connected text, research on reading comprehension for students with ID and ASD is limited (El Zein et al., 2016; Flores & Ganz, 2007). As few studies have examined reading comprehension interventions for students with ID, ASD and/or low IQ, little is known about best practices for improving reading comprehension in this population (Chiang & Lin, 2007; Flores & Ganz, 2007; O'Connor & Klein, 2004; Williamson et al., 2015). Researchers suggest that students with ID and ASD struggle to develop reading comprehension due to deficits in short-term memory, accessing prior knowledge, and deficiencies in oral language. These behaviors and others commonly associated with this population, such as difficulty with social interactions and self-stimulatory behaviors, can affect student responses to assessment, making it difficult to track growth over time (Jones et al., 2018; Ketterlin-Geller, 2008).

It is clear that additional research on comprehension interventions, as well as research on appropriate assessments that accurately measure reading comprehension behaviors, for students with ID and ASD is necessary. In order to conduct research on effective reading comprehension interventions for this population, researchers must use appropriate assessments that are sensitive to reading comprehension growth. Students who demonstrate deficits in reading comprehension often also demonstrate deficits in listening comprehension, engagement, and other behaviors that

are necessary, although not sufficient, to develop reading comprehension; these behaviors are often not captured on measures of reading comprehension (Leach et al., 2003; Yuill & Oakhill, 1991). Without adequate measures to assess either reading comprehension or the cascading levels of comprehension behavior, it is difficult to distinguish the specific intervention needs of struggling readers to support their growth in reading comprehension (Foorman et al., 2018).

The Systematic Observation of Language and Reading

The Systematic Observation of Language and Reading (SOLR) is a systematic observation tool that measures reading comprehension behaviors in beginning readers. Systematic observations are tools that can be used in classrooms to track student progress in academic content, inform instructional decisions, and hold educators accountable to specific standards (Wilhelm et al., 2018). The SOLR is a systematic observation tool that can be used by professionals, both practitioners and researchers, to measure student reading comprehension behaviors during literacy instruction. Additionally, the SOLR is inclusive of students with intensive needs, including students with disabilities such as ID and ASD.

The SOLR operationalizes reading comprehension through 25 specific behavior items that make up seven different constructs. The seven comprehension constructs that comprise the SOLR are: (a) language development, (b) abstract thinking, (c) elaboration, (d) print, (e) engagement, (f) fluency and prosody, and (g) off task/refusal behaviors. Each of the 25 behaviors are listed on the SOLR scoring sheet and described with examples in a handbook created by a team of researchers at Southern Methodist University (SMU) with expertise in behavior and reading research for students with disabilities. See Appendices A and B for examples of the scoring sheet and handbook. The behaviors are either observed (1) or not observed (0) by a rater at 30-second time intervals. Observations take place during literacy instruction and focus on one

student interacting with their teacher. The SOLR is designed to capture comprehension behaviors in one-on-one or small group settings, preferably from video recordings that can be paused at the discretion of the rater.

Summary and Purpose

The purpose of this paper is to examine the technical adequacy of the SOLR through estimates of interrater reliability and internal validity. In this paper, I describe the development of the SOLR, including the development of the behaviors listed on the SOLR and the comprehension constructs that are formed from combinations of the behaviors. The technical adequacy of the SOLR is examined through the following three research questions:

1. Is there strong overall interrater reliability on the SOLR?
2. Is the SOLR interrater reliability conditional on behavior?
3. Are the comprehension constructs listed on the SOLR adequately described by the behaviors?

Method

In this section, I discuss the development of the SOLR and the process for answering the three research questions. To measure interrater reliability and answer the first two research questions, I analyzed results from three independent raters using the SOLR to code the same set of videos, relying on the Gwet Agreement Coefficient (AC; Gwet 2014) as the interrater reliability statistic. Gwet AC estimates for each pair of raters are provided for each of the seven individual constructs, found in Table 1. Percent agreement, overall and between each construct, is found in Table 2. To measure construct validity and answer the third research question I conducted a survey on *Qualtrics* (Qualtrics, Provo, UT, 2005), a web-based software that allows the user to create polls and surveys, completed by teachers enrolled in master's level courses at

SMU. The teachers all had experience either teaching literacy or taking courses in the foundations of literacy instruction. The purpose of the assessment was to determine how well the seven constructs aligned with the behaviors listed on the SOLR. Images from each page of the *Qualtrics* (Qualtrics, Provo, UT, 2005) assessment are available in Appendix C.

Development of the SOLR

The SOLR was developed by a team of researchers with the goal of developing an observation tool that could measure comprehension behaviors in beginning readers and be sensitive enough to capture change in reading comprehension behavior for students with disabilities. During development of the SOLR, the researchers relied on the Systematic Assessment of Book Reading (SABR; Pentimonti et al., 2012), an observation tool that measures teacher behavior during literacy instruction. The SABR measures teacher behaviors that promote language and literacy development and is aligned to the national and state standards for students in prekindergarten to first grade (see <https://cliengage.org> for more information). The SABR includes two forms, a short form that is best used with video-recorded observations and a longer form that is used most effectively with transcriptions of read-aloud activities that take place between students and teachers. The SABR focuses on student responses to teacher asked questions during literacy instruction and tracks frequency of questions asked and comments made by students during instruction as well as teacher redirections and reminders.

The researchers refined the items listed on the SABR to capture student behavior and created additional behavior items necessary to capture seven specific constructs related to reading comprehension behaviors: (a) language development, (b) abstract thinking, (c) elaboration, (d) print, (e) engagement, (f) fluency and prosody, and (g) off task/refusal behaviors. The research team then met to discuss and further describe behaviors. During this discussion, the

raters used videos of students with ID and ASD engaging in literacy instruction during participation in a study examining the efficacy of a comprehensive reading curriculum designed for students with disabilities, *Friends on the Block* (FOTB; Allor, Cheatham, & Al Otaiba., 2018). The initial development of the SOLR was in direct response to the FOTB study. During the study, researchers gathered proximal and distal data that supported student growth in measures of sight and decodable word reading (see Allor et al., 2018, Allor et al., 2020; and friendsontheblock.com for more information). Student engagement and comprehension growth was observed throughout the study anecdotally by researchers, teachers, and parents. The researchers created the SOLR in an attempt to quantify the engagement and comprehension growth of students participating in the study.

Using video footage from students in the FOTB study, the researchers created a codebook that contained examples of each behavior listed on the SOLR to accompany the scoring sheet (see Appendices A and B). After refining the codebook, the raters reached consensus on the scoring of a 5-minute video of a student with ID participating in literacy instruction during the FOTB intervention. The behaviors were further refined and described in detail with the assistance of a trained graduate student during the rater agreement process, which all took place before individual interrater reliability was measured.

Interrater Reliability

To support the usefulness of the SOLR as an effective and accurate tool through which to measure reading comprehension behaviors, it is important to provide an estimate of interrater reliability to demonstrate that the measure can be used consistently (Springer, 2010). The interrater reliability of the SOLR is examined through the first two research questions, including examination of overall interrater reliability and interrater reliability among the seven constructs.

Interrater reliability is a way to quantify the closeness of scores provided by raters to the same participants of a study (Gwet, 2008). The ability to demonstrate reliability among raters allows us to infer that the data are reliable. Ideal agreement among raters on an observation tool is between 80-90 percent, with 70 percent also acceptable with “more complex instruments” (Wilhelm et al., 2018). Percent agreement is one way to calculate interrater reliability; however, certain tests of reliability, such as the Gwet AC, used in this study, produce a statistic for interrater reliability are often more appropriate (Gwet, 2008).

Data Collection

Data for interrater reliability was collected from three independent raters across videos of three students engaging in one-on-one literacy instruction. The video data was gathered from videos of student participants in a study of a comprehensive reading intervention (FOTB) created for students with disabilities, including ID and ASD (Allor, Cheatham, & Al Otaiba, 2018). Three students were selected for initial coding due to clarity of video (both visual and audio) and length (a significant portion of instruction spent on story reading). The first student, AMM, was a 9-year-old boy with a diagnosis of ASD and an IQ of 42. The second student, CF, was a 9-year-old girl with Down syndrome and an IQ of 42. The third student, EM, was a 6-year-old girl with a diagnosis of Smith-Lemli Opitz syndrome and an IQ of 59. Additional demographic information about the students can be found in Table 3. Each participant attended a private school for children with disabilities in Texas and was recruited for the study due to deficits in reading development. To calculate reliability, each of the three coders, two professors and one graduate student, all of whom had research expertise in special education and experience as K-12 special education teachers, coded one video per student. Videos, which ranged in length from six to eight minutes, captured one-on-one literacy instruction between the student and teacher during

the book-reading portion of the FOTB intervention (see www.friendsontheblock for more information). The book-reading portion of the FOTB lesson was designed to promote student-teacher interaction through turn taking and dialogic reading questions, which resulted in an opportunity for many of the SOLR behaviors to be exhibited across participant videos. Table 4 provides a count of the observed behaviors in each of the seven constructs by the individual raters across the three students.

Gwet AC Statistic

Upon examining the individual rating sheets from the videos coded for reliability, I recognized that scores from the raters reflected a high prevalence of agreement. At times, when the prevalence of agreement is high, meaning there is a large extent of agreement between raters, this high agreement is not reflected in the test statistic (Gwet, 2014). I examined several test statistics to determine which would be the most appropriate, and ultimately decided to use the Gwet AC (Gwet, 2008). The Gwet AC is an agreement coefficient used to estimate interrater reliability, and when rater agreement is high, the Gwet AC is an appropriate coefficient to use for computing interrater reliability, as other coefficients are often unreliable (Gwet, 2008). The assumptions met when using the Gwet AC to calculate reliability are “(a) Chance agreement occurs when at least one rater rates an individual randomly and (b) Only an unknown portion of the observed ratings is subject to randomness.” (Gwet, 2008, p. 35). Gwet also describes the prevalence for biased estimates when relying upon common coefficients and these biases are not reflected in her AC statistic, as the code and assumptions are designed to estimate true interrater reliability under extreme circumstances (Gwet, 2008).

Internal Validity

“Validity is the most fundamental consideration in developing and evaluating tests – the process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretation” (AERA, APA, & NCME, 1999, p. 9). Test sufficiency, clarity, relevance, and the match between the items and tasks or constructs are all integral pieces of internal validity information (Goodwin & Leech, 2003). To examine the internal validity of the SOLR and describe the extent to which the constructs are accurately defined by the behaviors, I created a matching assessment on *Qualtrics* (Qualtrics, Provo, UT, 2005), which was completed by teachers enrolled in masters level literacy courses at SMU. One hundred three participants completed the survey that served as the measure of internal validity and answered research question three. A majority of the participants were in at least the second semester of an education master’s program. The survey took an average of 15 minutes to complete. All participants were provided time during class to complete the survey and were ensured that participation was both anonymous and voluntary. The ultimate goal of the SOLR is to create a comprehension measure that can be utilized not only by researchers but also by classroom teachers to inform instruction. Therefore, the master’s level teachers were appropriate candidates to complete the survey assessment to capture internal construct validity, as their level of understanding would likely inform the feasibility of the tool for use with practitioners.

The survey was designed to answer the third research question, which examines the extent to which the SOLR constructs are accurately defined by the behaviors. The first page of the survey contained information regarding informed consent. On the second page of the survey, the teachers were provided with a list of the 25 behaviors on the SOLR, along with seven boxes containing the comprehension constructs supported with a short definition of the construct. The seven constructs and their definitions as listed on the assessment are: 1) language development –

the ability to express language; 2) abstract thinking – higher level cognitive processes; 3) elaborations – the ability to make connections between what the student knows and what is present in the book; 4) print – understanding that print has form and functions; 5) engagement – attending to the book and pictures; 6) fluency/prosody – reading with automaticity, accuracy, and expression; 7) off task/refusal – student exhibits behaviors that interfere with ability to learn. Participants were instructed to drag each behavior into the box containing the construct to which they believed the behavior belonged and were required to match each behavior to a construct before submitting their responses.

Results

Research Question 1: Is There Strong Overall Interrater Reliability on the SOLR?

According to the estimates of interrater reliability, using the Gwet AC, interrater reliability on the SOLR is strong overall. The Gwet AC only compares pairs of raters, not three raters at a time, so I have listed specific estimates for each pair of raters. Gwet AC estimates with 95% confidence bounds for the pairs of raters are; raters 1 and 2, $AC_1=0.90$ (0.88, 0.92); raters 2 and 3, $AC_1=0.91$ (0.89, 0.93); and raters 1 and 3, $AC_1=0.87$ (0.85, 0.90). This information, including confidence bounds, for each pair of raters can also be found in Table 1.

Research Question 2: Is the SOLR Interrater Reliability Conditional on Behavior?

Results from the Gwet AC analyses, which produce an estimate for interrater reliability between the raters for each of the seven constructs, support that interrater reliability is conditional on behavior. For constructs such as abstract thinking, the Gwet AC estimate was as high as $AC_1=0.99$ (0.98, 1.00) (raters 1 and 2) and $AC_1=0.99$ (0.97, 1.00) (raters 2 and 3). The lowest estimate was engagement for raters 1 and 3, at $AC_1=0.56$ (0.41, 0.71). The variability in estimate among constructs is provided in Table 1, with particular attention to the behaviors of

engagement and fluency/prosody, which were consistently among the lowest reliability estimates for each pair of raters. However, the issue remains as to whether or not these differences in interrater reliability among constructs are statistically significant.

To determine if the differences in rater reliability among construct were statistically significant, I examined the confidence bounds from each Gwet AC estimate, provided in Table 1. The confidence bounds serve as a proxy for statistical significance, and most estimates of agreement are close to 0.90 or include 0.90 within the confidence bounds, indicating high interrater reliability. The two constructs for which the confidence bounds do not include the estimate of 0.90 (among any pair of raters) are the constructs of engagement and fluency/prosody. The three Gwet AC estimates for the construct of engagement were $AC_1=0.72$ (0.60, 0.85) (raters 1 and 2); $AC_1=0.77$ (0.66, 0.89) (raters 2 and 3); and $AC_1=0.56$ (0.41, 0.71) (raters 1 and 3). The three Gwet AC estimates for the construct of fluency/prosody were $AC_1=0.77$ (0.67, 0.87) (raters 1 and 2); $AC_1=0.77$ (0.68, 0.87) (raters 2 and 3); and $AC_1=0.80$ (0.71, 0.89) (raters 1 and 3). Additionally, the interrater reliability estimates provided by the Gwet AC for the constructs of language development, abstract thinking, and elaborations are consistently above 0.94 among each of the three pairs of raters.

Research Question 3: Are the Constructs Listed on the SOLR Adequately Described by the Behaviors?

Results from the survey used to address research question three, which examines whether the constructs listed on the SOLR are adequately described by the behaviors, support that the constructs are adequately described by the behaviors. However, some constructs are better described by the behaviors than others are, and the percentage that each item was matched to the construct it was assigned to on the SOLR is provided in Table 4. Overall, the participants

matched behaviors to a construct with a relatively high rate of agreement to what is listed on the SOLR. The rate of agreement was as follows, from highest to lowest: off task/refusal (98%), engagement (75%), abstract thinking (74%), fluency (72%), elaborations (58%), language development (57%), and print (27%).

In addition to percent accuracy of each overall construct, Table 4 also provides information at the item level. There were several behaviors which were matched with very high accuracy to their SOLR construct, for example both behaviors in the off task/refusal construct (100% and 96%), as well as the following four behaviors: *“reads words accurately or with minimal errors and/or prompting”* (85%), *“demonstrates excitement/engagement about story, character, or actions in the book”* (84%), *“demonstrates excitement/engagement about reading the illustrations/pictures”* (83%), and *“uses pitch, stress, intonation to convey meaning”* (83%). However, several items were not matched to their SOLR construct with high accuracy. For example, *“initiates reading without prompting”* (12%), *“is involved with turn taking with text reading-verbalizing”* (15%), and *“extends conversation with relevant questions or story remarks”* (32%) are a few items the participants did not match with high accuracy.

Results from the survey show a wide range of SOLR construct choices for seven of the specific behaviors listed on the tool. For these seven behaviors, participants matched the behavior to the construct listed on the SOLR with less than 50% accuracy, instead indicating that the behaviors were a better fit with other constructs listed on the SOLR. Information about these seven behavior items is presented in Table 6, in which the percent to which the participants matched these behaviors to each of the seven SOLR constructs is provided. For several of the behaviors that were matched with less than 50% accuracy, a larger number of participants agreed that the behavior belonged to another construct listed on the SOLR, as opposed to the SOLR

construct the behavior belonged. For example, on the SOLR, the behavior “*extends conversation with relevant question or story remarks*” was aligned to the construct of elaborations. According to the survey, 32% of the participants believed this behavior belonged to the construct of elaborations, however another 33% indicated they believed the behavior belonged in language development, and another 33% indicated they believed it belonged to abstract thinking. Another behavior, “*is involved in turn taking with text reading-verbalizing*” was matched to its SOLR construct of print with 15% accuracy; however, 33% of participants matched this behavior to language development and another 36% to engagement. Finally, the behavior “*initiates reading without prompting*” was matched to its SOLR construct of print with 12% accuracy, and a majority of participants, 53%, indicated they believed this behavior belonged to the construct of engagement.

Discussion

The purpose of this paper was to evaluate the adequacy of an observation tool, the SOLR, designed to capture reading comprehension behaviors of students with ID and ASD through research questions regarding interrater reliability and test validity. Findings support strong interrater reliability, both overall and among construct, with the strength of the interrater reliability conditional on behavior. Findings also support strong construct validity, as the behaviors listed on the SOLR were adequately aligned to their constructs, with some variability among behavior. Interrater reliability and test validity are important to consider when determining the technical adequacy of assessment tools. Interrater reliability is a common method to quantify the closeness of scores given to the same participants of a study and assures that multiple raters would come to similar conclusions when observing the same participant (Gwet, 2008). Test validity is also important to consider when designing assessment tools, as

validity provides a basis for score interpretations and sufficiency of constructs captured on the assessment. In this manuscript, I have described the process for obtaining both interrater reliability and test validity, as well as the results. Findings suggest that strong overall interrater reliability on the SOLR was achieved, but that the interrater reliability was conditional on behavior construct. Findings also support strong test validity by examining the extent to which the behaviors on the SOLR adequately describe their construct. The results presented in this manuscript are somewhat promising; however, there are also several issues to discuss related to the research questions.

Research Question One

The first research question asks if there is strong overall interrater reliability on the SOLR. Results from percent agreement and the Gwet AC support strong overall interrater reliability. I chose to use the Gwet AC (Gwet, 2008) for statistical analysis of interrater reliability, as the underlying assumptions when using the Gwet AC consider the presence of high agreement, which was present, as percent agreement was 89.7%. The Gwet AC provides an estimate that is representative of the high percent agreement among raters; however, this estimate compares only two raters at a time, and reliability information was obtained from three individual raters. As seen in Table 1, Gwet AC estimates are provided for each pair of raters, which were high for each pair of raters, indicating strong interrater reliability on the SOLR. The estimates for each pair of raters, overall, are listed here and also outlined in Table 1: raters 1 and 2: $AC_1 = 0.90$ (0.88, 0.92); raters 2 and 3: $AC_1 = 0.91$ (0.89, 0.93); raters 1 and 3: $AC_1 = 0.87$ (0.85, 0.90).

Research Question Two

Through the second research question, I examine whether interrater reliability is conditional on behavior by examining Gwet AC estimates for each pair of raters across the seven constructs measured by the SOLR, with emphasis on their confidence bounds. Results indicated that interrater reliability was, in fact, conditional on construct. For some constructs, such as language development, abstract thinking, and elaborations, raters were overall more reliable than on several other behaviors, with the lowest reliability consistently among the constructs of engagement and fluency/prosody (see Table 1 for the individual Gwet AC estimates by construct).

It is possible that additional rater training on the constructs with the lowest interrater reliability would help to improve interrater reliability estimates for these constructs. More emphasis could be placed on the behaviors in the constructs for which raters were less reliable during the training phase to improve later interrater reliability in these areas. For instance, engagement (attending to the book and pictures) and fluency/prosody (reading with automaticity, accuracy, and expression) are two constructs for which the raters were less reliable. During the rater agreement process, the raters discussed the behaviors in these constructs more than others while watching student videos, as there was more subjectivity to whether or not the behaviors were observed. Additional rater training, focusing on these behaviors, could improve rater reliability in these constructs in the future.

Additionally, interrater reliability for the constructs of engagement and fluency/prosody could possibly be improved if the team were to revisit the definitions of the behaviors in these constructs and rephrase the items. For example, one of the behaviors that belongs to the construct of engagement, “*physically engages with the book by holding it, pointing within it, and visually attending to it*” could be rewritten to be more specific or have examples of this behavior

expanded upon in the codebook. This behavior was one that participants of the survey struggled to match the behavior to the construct of engagement as compared to the other behaviors that belonged to this construct.

There is also the possibility that these two constructs, engagement and fluency/prosody, are constructs that are more difficult to capture through observation than the other constructs described on the SOLR, particularly when observing students with ID and ASD. Students with ID and ASD may demonstrate behaviors that make it difficult to determine whether or not they are truly engaged/speaking with fluency, such as task avoidance, limited eye contact, echolalia, and deficits with receptive language. For example, one student whose videos were used for interrater reliability, AMM, was a student with ASD who displayed echolalia behaviors during reading instruction, meaning he would often repeat things his teacher said to him later during the lesson. Because of this behavior, it was often difficult to tell if he was actually reading with fluency or simply imitating what his teacher previously said to him. Additionally, CF was a student with Down syndrome who displayed many task-avoidance type behaviors when frustrated or when working one-on-one. For example, she often avoided making direct eye contact with the book or the teacher, making it difficult to determine if she was engaged with the reading. To thoroughly examine whether or not differences in rater reliability among constructs is due to the descriptions of the behaviors in the constructs or the behaviors often exhibited by students with ID and ASD, I examined the results of the survey used to answer research question three, discussed next.

Research Question Three

The third research question examines how well the constructs listed on the SOLR are described by the behaviors to which they are aligned. Results indicate, overall, behaviors

described their construct adequately, with some variability among construct. To answer this research question, a survey was completed by students enrolled in master's level education courses at SMU. Many of the participants were practicing teachers, and all had either experience teaching foundational literacy courses or participated in literacy instruction courses at SMU. Results from the 103 participants who completed the survey are presented in Tables 4 and 5.

Overall, the results from the survey suggest that behaviors were aligned to the constructs, with varying degrees of strength. I expected that the constructs with the least percentage of correctly matched items would be engagement and fluency/prosody, the same two constructs for which raters struggled with reliability and agreement. However, the construct of engagement was one of the constructs for which the participants matched each behavior to the construct with higher accuracy than many of the other behaviors, at 75% accuracy overall. Similarly, the construct of fluency/prosody was 72% accurately matched overall, which was high compared to other constructs (see Table 6). Overall, the construct of off task/refusal was the highest correctly matched construct, at 98%, followed by engagement (75%), abstract thinking (74%), and fluency/prosody (72%). The constructs matched with the least accuracy were print (27%), language development (57%), and elaborations (58%).

These results from the survey suggest that, just as interrater reliability was conditional on construct, description of construct by behaviors was also conditional on construct. However, the difference was that the more accurately defined constructs (i.e., constructs with higher validity) were not the constructs that the raters achieved the highest interrater reliability (examined through research question two). For example, the construct of elaborations, for which interrater estimates were consistently high between raters, was only matched with 58% accuracy to its construct listed on the SOLR. One behavior that belonged to the construct of elaborations,

“extends conversation with relevant question or story remarks” was matched by 33% of participants into the construct of language development and another 33% of participants into the construct of abstract thinking. Another construct for which the participants struggled to identify behaviors from the SOLR was the construct of print; however, print was not a construct of concern during interrater reliability analyses. One behavior for the construct of print, *“initiates reading without prompting”*, was only correctly matched by 12% of participants. A majority of participants, 53% placed this behavior into the construct of engagement. Another behavior intended for the construct of print, *“is involved in turn taking with text reading-verbalizing”*, was only correctly matched by 15% of participants, with 33% of participants placing this behavior into language development and another 36% into engagement. See Table 6 for information regarding each of the seven items that were matched to their construct listed on the SOLR at less than 50% accuracy, including the percent to which those behaviors were matched to each of the SOLR constructs.

The results of the survey support three constructs for which the participants matched behaviors to their intended construct with the least accuracy: print, language development, and elaborations. While these three constructs were not the constructs which the raters struggled most to reach interrater reliability, the participants did match several of the behaviors that belonged to print, elaborations, or language development, into the construct of engagement, a construct for which the raters achieved less interrater reliability. It is possible that due to the lack of clarity around the construct of engagement and general difficulty identifying the behaviors that are used to operationalize construct by both researchers and practitioners, the survey participants found that many of the behaviors that were intended to represent other constructs also represented behaviors that describe engagement.

Before the SOLR is disseminated to a larger network of researchers or practitioners, the behaviors that the survey participants believed described a different construct than currently described on the SOLR should be reviewed in light of these findings. Based on results from the survey, these behaviors could be either rewritten or placed into a different behavior construct. However, it is possible that if the survey were conducted with a different group of participants, such as researchers or special education teachers, results could indicate that the behaviors do not need to be revised. Additionally, training on the SOLR or on the constructs listed could also result in behaviors being matched to their intended construct at a higher rate. Recommendations for improving the SOLR based on survey results, as well as implications for researchers and practitioners, are discussed further in the following section.

Limitations and Directions for Future Research

Several limitations must be addressed with respect to the design of this study, beginning with limitations surrounding data analysis for interrater reliability. Unfortunately, the issue of low estimate of interrater reliability in presence of high rater agreement is a common issue. However, use of the Gwet AC (Gwet, 2008) adequately recognized the presence of high agreement through the underlying assumptions and was the best fit to analyze this data.

Other issues with interrater reliability surround the collection of the data itself, as this study was designed after the FOTB (Allor, Cheatham & Al Otaiba, 2018) study had been completed. The videos were taken for fidelity purposes, to ensure that the teachers were implementing the curriculum in the manner that it was designed and not for the purposes of capturing reading comprehension behaviors. As the purpose of the videos was not to track comprehension growth, the interactions of students and teachers vary throughout the videos, providing different amount of opportunities to respond throughout different videos. At times, the

differences in opportunities to respond were directly related to the day of the week that the video was recorded, as lesson structure changed daily and focused on skills such as predicting, reading fluency, or answering dialogic reading questions. Due to this limitation, I made the best choices with the videos that were available for data collection. I attempted to search for video data that contained an adequate amount of student and teacher interactions so that opportunities to exhibit most, if not all, of the behaviors described on the SOLR were present at some point throughout the lesson.

Another limitation related to the development of the SOLR is that the reliability and validity of the tool were collected simultaneously, instead of collecting validity information prior to obtaining interrater reliability data. The team of researchers began using the SOLR to code videos for interrater reliability before receiving results from the construct validity assessment, the survey. Results from the survey suggested the possibility of editing some of the behaviors listed on the SOLR, particularly the items that the survey participants had the most trouble correctly matching to their SOLR construct. Ideally, the team of researchers would have developed the tool and then conducted the survey so that behaviors and descriptions could be altered before beginning to code videos for interrater reliability.

Another limitation of this study involves the sample that was chosen to complete the survey. I chose to use students enrolled in master's level education courses for two reasons. One reason is that the goal of the SOLR is to have a tool that will be used not only by researchers, but also by practitioners. The second reason is that teachers enrolled in master's level programs involving literacy instruction would likely have sufficient background knowledge in the concepts of reading comprehension to respond meaningfully to the survey. However, this sample did not include researchers or special education teachers. Although the chosen sample of teachers was

appropriate and readily available, a broader sample of teachers with a wider array of literacy experience might have produced different results. Future research on the development of the SOLR could involve the same survey but delivered to education researchers for comparison, especially as the SOLR was designed by and is currently only in use by education researchers. It is also possible that practicing special education teachers would have yielded different results as special education teachers often receive different training than general education teachers during masters and undergraduate programs.

Implications for Practice

The SOLR is appropriate for researchers to use when measuring reading comprehension behaviors of beginning readers, including students with ID and ASD, as it is designed to be used most effectively with video-recorded student teacher interactions which are relatively easy to collect and do not require real time coding. However, in its current form, the SOLR may not be appropriate for practitioners to use as a tool to progress monitor or assess growth, as a more streamlined measure that could be completed while conducting a lesson or directly after a lesson would be more feasible. Examples include a checklist version of the SOLR that teachers could use during instruction, or shortened versions in which the teacher focuses on tracking only one construct at a time.

Future paths of inquiry should focus on how the SOLR could potentially be used by teachers and administrators to progress monitor. A shorter form, similar to a checklist, could be developed and would likely be more feasible for practitioner use than the current form of the SOLR, as checklists are utilized by special education practitioners in the classroom when tracking behavior. Again, a checklist form, or condensed version of the SOLR, might include only one construct at a time. The teacher could choose to focus on a construct that the student

might have a weakness in, or begin by targeting the construct that would be the simplest to master before moving to the more difficult constructs. Implementing one checklist per construct in real time is feasible, as the constructs contain at most five behaviors to track.

Conclusion

Interrater reliability, overall, and even among the constructs for which raters were less reliable, of the SOLR was strong. Additionally, the results from the survey examining the validity of the SOLR were promising, providing insight into a few behaviors that were aligned to their intended construct with more strength than others. Ultimately, the results from these analyses suggest the SOLR can be used by researchers to adequately monitor comprehension behaviors exhibited by students with ID and ASD. The SOLR has potential to be used by practitioners, such as classroom teachers, to measure comprehension behaviors, though a detailed process for training should be set in place and a more feasible form should be developed. Additionally, greater explanation of the specific behaviors, with examples, should be added into the codebook so that users will have a clear understanding of what specific comprehension behaviors might look like when exhibited by students with ID and ASD. The SOLR is an important tool to continue to refine in order to provide an observation tool for assessing comprehension behaviors of beginning readers, including readers with disabilities.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). Standards for educational and psychological counseling. Washington, DC: American Educational Research Association.
- Allor, J. H., Cheatham, J., & Al Otaiba, S. (2018). *Friends on the block*. Dallas, TX: Friends on the Block.
- Allor, J. H., Gifford, D. B., Jones, F. G., Al Otaiba, S., Yovanoff, P., Ortiz, M. B., & Cheatham, J. P. (2018). The effects of a text-centered literacy curriculum for students with intellectual disability. *American Journal for Intellectual and Developmental Disabilities, 123*, 474-494. doi:10.1352/1944-7558-123.5.474
- Allor, J. H., Mathes, P. G., Roberts, J. K., Cheatham, J. P., & Otaiba, S. A. (2014). Is scientifically based reading instruction effective for students with below-average IQs?. *Exceptional Children, 80*, 287-306.
- Allor, J., Yovanoff, P., Al Otaiba, S., Ortiz, M., & Conner, C. (2020). Evidence for a literacy intervention for students with intellectual and developmental disabilities. *Education and Training in Autism and Developmental Disabilities, 55*, 290-302.
- Browder, D. M., Ahlgrim-Delzell, L., Courtade, O., Gibbs, S. L., & Flowers, C., (2008). Evaluation and the effectiveness of an early literacy program for students with significant developmental disabilities. *Exceptional Children, 75*, 33-52.
- Browder, D. M., & Spooner, F. (2014). *More language arts, math, and science for students with severe disabilities*. Baltimore, MD: Brookes.

- Chiang, H., & Lin, Y. (2007). Reading comprehension instruction for students with autism spectrum disorders: A review of the literature. *Focus on Autism and Other Developmental Disabilities, 22*, 259-267.
- Connor, C. M., Alberto, P. A., Compton, D. L., & O'Connor, R. E. (2014). Improving Reading Outcomes for Students with or at Risk for Reading Disabilities: What Have We Learned? *Synthesis of IES Research*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277-299.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.
- Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*, 561-584.
- El Zein, F., Solis, M., Lang, R., & Kim, M. K. (2016). Embedding perseverative interest of a child with autism in text may result in improved reading comprehension: A pilot study. *Developmental Neurorehabilitation, 19*, 141-145.
doi:10.3109/17518423.2014.915893ESSA, 2015
- ESSA (2015). Every Student Succeeds Act of 2015, Pub. L. No. 114-95 § 114 Stat. 1177 (2015-2016).
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10*, 323-330.

- Flores, M. M., & Ganz, J. B. (2007). Effectiveness of direct instruction for teaching statement inference, use of facts, and analogies to students with developmental disabilities and reading delays. *Focus on Autism and Other Developmental Disabilities*, 22, 244-251.
doi:10.1177/10883576070220040601
- Foorman, B. R., Arndt, E. J., & Crawford, E. C. (2011). Important constructs in literacy learning across disciplines. *Topics in Language Disorders*, 31, 73-83.
doi:10.1097/TLD.0b013e31820a0b86
- Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading*, 10, 301-322.
- Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development*, 36, 181-191.
[doi:10.1080/07481756.2003.11909741](https://doi.org/10.1080/07481756.2003.11909741)
- Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, (1), 29-48.
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC.
- Jones, F. G., Gifford, D., Yovanoff, P., Al Otaiba, S., Levy, D., & Allor, J. (2019). Alternate assessment formats for progress monitoring students with Intellectual Disabilities and below average IQ: An exploratory study. *Focus on Autism and Other Developmental Disabilities*, 34, 41-51.

- Katims, D. S. (2001). Literacy assessment of students with mental retardation: An exploratory investigation. *Education and Training in Mental Retardation and Developmental Disabilities*, 363-372.
- Ketterlin-Geller, L. R. (2008). Testing students with special needs: A model for understanding the interaction between assessment and student characteristics in a universally designed environment. *Educational Measurement: Issues and Practices*, 27, 3–16.
- Knott, R., Patterson, K., & Hodges, J. R. (1997). Lexical and semantic binding effects in short-term memory: Evidence from semantic dementia. *Cognitive Neuropsychology*, 14, 1165–1216.
- Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology*, 95, 211. doi:10.1037/0022-0663.95.2.211
- Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child psychology*, 73, 139-158. doi: [10.1006/jecp.1999.2498](https://doi.org/10.1006/jecp.1999.2498)
- No Child Left Behind Act of 2001, P.L. 107-101, 20 U.S.C. § 6319 (2002).
- O'Connor, I. M., & Klein, P. D. (2004). Exploration of strategies for facilitating the reading comprehension of high-functioning students with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 34, 115-127. doi:10.1023/B:JADD.0000022603.44077.6b
- Pentimonti, J. M., Zucker, T. A., Justice, L. M., Petscher, Y., Piasta, S. B., & Kaderavek, J. N. (2012). A standardized tool for assessing the quality of classroom-based shared reading:

- Systematic Assessment of Book Reading (SABR). *Early Childhood Research Quarterly*, 27, 512-528.
- Qualtrics, Provo, UT, 2005
- Springer, K. (2010). *Educational research: A contextual approach*. Hoboken, NJ: Wiley.
- Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. *Reading and Writing*, 4, 245–256.
- Wilhelm, A. G., Rouse, A. G., & Jones, F. (2018). Exploring Differences in Measurement and Reporting of Classroom Observation Inter-Rater Reliability. *Practical Assessment, Research & Evaluation*, (4).
- Williamson, P., Carnahan, C. R., Birri, N., & Swoboda, C. (2015). Improving comprehension of narrative using character event maps for high school students with autism spectrum disorder. *The Journal of Special Education*, 49, 28-38. doi:10.1177/0022466914521301
- Yuill, N., & Oakhill, J. V. (1991). *Children's problems in text comprehension*. Cambridge, UK: Cambridge Univ. Press

Table 1*Gwet AC: rater 1 and rater 2*

Pair of Raters	Behavior category	Estimate	Standard error	Lower CB	Upper CB
Rater 1 and Rater 2	1	.94	.02	.90	0.97
	2	.99	.01	.98	1.01
	3	.98	.01	.96	1.00
	4	.79	.06	.68	0.90
	5	.72	.06	.60	0.85
	6	.77	.05	.67	0.87
	7	.96	.03	.90	1.01
	overall	.90	.01	.88	0.92
Rater 2 and Rater 3	1	.95	.02	.92	0.99
	2	.99	.01	.97	1.01
	3	.94	.02	.90	0.98
	4	.95	.03	.90	1.01
	5	.77	.06	.66	0.89
	6	.77	.05	.68	0.87
	7	.93	.03	.86	0.99
	overall	.91	.01	.89	0.93
Rater 1 and Rater 3	1	.95	.02	.92	0.99
	2	.98	.01	.96	1.00
	3	.95	.02	.91	0.98
	4	.76	.06	.66	0.89
	5	.56	.07	.41	0.71
	6	.80	.05	.71	0.89
	7	.88	.04	.80	0.97
	overall	.87	.01	.85	0.90

Table 2*Percent agreement: all raters*

Behavior	Percent agreement
Overall	89.7
Language Development	93.7
Abstract Thinking	98.2
Elaborations	93.9
Print	87.8
Engagement	76.4
Fluency/Prosody	82.9
Off Task Behavior	90.2

Table 3*Student demographic information*

Case	Age	IQ	PPVT age eq	Disability	Total dosage (hours)
AMM	9	42	2:06	ASD	26.48
CF	9	42	2:09	Downs syndrome	30.46
EM	9	59	2:04	Smith-Lemli Opitz	31.96

Table 4*Count of Observed Behaviors by Rater*

Participant	Count of Observed Behaviors by Construct						
	Language Development	Abstract Thinking	Elaborations	Print	Engagement	Fluency/Prosody	Off Task/ refusal
AM							
rater 1	2	1	0	20	14	20	0
rater 2	2	1	0	20	14	19	0
rater 3	5	0	0	22	12	18	0
CF							
rater 1	6	0	1	25	18	22	4
rater 2	1	1	3	24	15	20	2
rater 3	1	0	0	27	24	18	7
EM							
rater 1	16	1	5	26	27	25	3
rater 2	17	2	4	28	23	26	4
rater 3	17	1	7	19	31	24	3

Table 5*Qualtrics matching assessment results*

Behavior	SOLR construct	% correctly matched	
		item	construct
Responds using a complete and meaningful thought	Language development	74	57
Labels or describes story actions		38	
Labels or describes nouns, locations, characters		56	
Provides descriptions/characteristics		40	
Gives a definition for a word		77	
Is able to compare or contrast	Abstract thinking	78	74
Offers an understanding from the character's point of view		68	
Makes a prediction or hypothesis about future events		76	
Offers an inference, analysis, or explanation	Elaborations	74	
Makes a text to life connection		75	58
Describes/recalls information about text from prior pages or previous readings		45	
Extends conversation with relevant question or story remarks		32	
Makes a text to text connection between books read	Print	79	
Is involved in turn taking with text reading-verbalizing		15	27
Demonstrates understanding of one-to-one correspondence		55	
Initiates reading without prompting	Engagement	12	
Demonstrates excitement/engagement about reading the illustrations/pictures		83	75
Demonstrates excitement/engagement about story, character or actions in the book		84	
Physically engages with the book by holding it, pointing within it, and visually attending to it		58	
Reads in meaningful phrases	Fluency, prosody	73	72
Reads words accurately or with minimal errors and/or prompting		85	
Student independently attempts to self-correct, repeat to problem solve, and/or sound out words		48	
Uses pitch, stress, and intonation to convey meaning	Off task/refusal	83	
Refuses to participate in reading (e.g. verbalizes refusal, puts head on desk, tried to escape)		100	98
Visually disengages with text or task by looking away from teacher or book		96	

Table 6*Qualtrics matching assessment: Behaviors matched below 50%*

Behavior	Language development	Abstract thinking	Elaborations	Print	Engagement	Fluency/prosody
Labels or describes story actions	38%*	9%	17%	23%	12%	1%
Provides descriptions/characteristics	40%*	15%	22%	11%	13%	0%
Describes/recalls information about text from prior pages or previous readings	28%	20%	45%*	2%	4%	0%
Extends conversation with relevant question or story remarks	33%	33%	32%*	1%	1%	0%
Is involved in turn taking with text reading-verbalizing	33%	7%	2%	15%*	36%	8%
Initiates reading without prompting	14%	5%	4%	12%*	53%	13%
Student independently attempts to self-correct, repeat to problem solve, and/or sound out words	17%	7%	4%	14%	11%	48%*

Chapter IV: Exploring the Sensitivity of the Systematic Observation of Language and Reading across Students with Intellectual Disability and Autism Spectrum Disorder

Students with Intellectual Disability (ID) and Autism Spectrum Disorder (ASD) have historically had less opportunity to engage in reading instruction beyond the use of functional sight words, compared to their general education peers, who receive comprehensive reading instruction that includes phonics and gaining meaning from print (Conners, 1992; Katims, 2000). More recent research documents students with intensive needs, such as ID and ASD, respond positively to comprehensive reading instruction that includes phonics instruction; however, continued research on the most effective reading comprehension interventions for this population is needed (Allor et al., 2014; Browder et al., 2008; Connor et al., 2014). In addition to continued research on effective reading interventions, there remain questions around how to most effectively measure reading skills for students with disabilities (Lemons & Fuchs, 2010). Adequate comprehension measures are needed to assess student progress and support their growth; however, many measures of reading comprehension behavior do not address other types of behavior that are necessary to develop comprehension, such as engagement, oral language, or listening comprehension (Cutting & Scarborough, 2006; Foorman et al., 2018; Francis et al., 2006; Yuill & Oakhill, 1991).

Reading comprehension measures are needed that accurately and sensitively measure comprehension behaviors and are also inclusive of beginning readers, with and without disabilities. Many assessments of reading comprehension involve listening tasks, demonstrating the ability to memorize the final/first word of a sentence, or reading a series of sentences and answering literal recall questions (Daneman & Carpenter, 1980, 1983; Leach et al., 2003; Fletcher, 2006; Nation et al., 1999; Stothard & Hulme, 1992;). It is necessary to develop

accurate and valid measures of reading comprehension, as assessments of reading comprehension are used to capture student abilities and can be used to inform instruction (Cutting & Scarborough, 2006; Fletcher, 2006; Francis et al., 2006). However, questions remain regarding how to most accurately and effectively assess growth in literacy skills for students with disabilities, who often display many individual differences in their response to intervention and delayed growth on curriculum-based progress monitoring as compared to their same aged peers (Lemons & Fuchs, 2010; Jones et al., 2018). Many extraneous variables, such as issues with attention span, oral language weaknesses, and sensory and behavioral issues, are often associated with students with disabilities such as ID and ASD, which can affect response to assessments and the technical adequacy of assessments (Ketterlin-Geller, 2008).

The Systematic Observation of Literacy Engagement (SOLR) is an observation tool that measures reading comprehension behaviors for beginning readers and is inclusive of students with intensive needs, such as ID and ASD. Developed by a team of special education researchers, the SOLR is a comprehension measure that was designed to accurately and sensitively assess growth of students with disabilities. SOLR scores can be used to determine response to reading instruction and inform instructional decisions. The SOLR, best used with video recording of student-teacher interactions, contains 25 behaviors that are either observed or not observed by a rater at 30-second intervals. The 25 behaviors make up seven constructs of comprehension behavior, which are language development, abstract thinking, elaborations, print, engagement, fluency/prosody, and off task/refusal behavior.

Reliability and validity are crucial in test development and should be established before the SOLR is used by researchers or practitioners to track comprehension growth over time. Results from Conner's (2020b) examination of the technical adequacy of the SOLR find that the

observation tool is both reliable and internally valid. To determine reliability, Conner (2020b) examined interrater reliability gathered from SOLR scores obtained by three raters across several videos of students with ID and ASD engaging in literacy instruction. According to the findings, raters reached a high level of reliability across all videos, although interrater reliability did vary across construct; however, reliability was still high across all seven constructs. Additionally, Conner (2020b) examined the internal validity of the SOLR through a survey in which students enrolled in masters level education courses were asked to match each of the 25 behaviors to the corresponding comprehension. Again, results suggest that overall, the descriptions of the behaviors listed on the SOLR are valid, and although strength of validity varied slightly among construct, validity was adequate across all constructs (Conner, 2020b). As the SOLR has been found to be both reliable and internally valid, the subsequent step in test validation is to document that the tool is sensitive to change over time.

The SOLR is likely sensitive to growth, particularly with students with disabilities, as the measure includes a wide range of constructs from a cascade of reading comprehension skills. Students ID and ASD often have deficits in oral language, accessing prior knowledge, and navigating social interactions; these deficits and externalizing behaviors can affect response to assessment (Jones et al., 2018; Ketterlin-Geller, 2008). Many current approaches to measurement of comprehension are relatively narrow and do not reflect the multidimensional nature of reading comprehension (Fletcher, 2008). It is possible that the SOLR, an observational tool, is likely to capture the multiple aspects of reading comprehension within the seven different constructs through which behaviors are observed.

Purpose

The purpose of this paper is to demonstrate that the SOLR is sensitive to comprehension growth over time for students with ID, ASD, and low IQ. I examine the sensitivity of the SOLR through video data of a group of students with disabilities involved in an initial efficacy study for the *Friends on the Block* (FOTB; Allor, Cheatham, & Al Otaiba, 2018; Allor et al., 2018) intervention over the course of one academic school year. The FOTB intervention is a comprehensive, researcher created intervention designed for students with disabilities, including those with ID, ASD, and low IQ (Allor et al., 2018; Allor et al., 2020). During the study examining the initial efficacy of the FOTB intervention, research assistants recorded video data of students and teachers engaging in instruction approximately once per month (Allor et al., 2018; Allor et al., 2020). In this manuscript, I examine whether the SOLR is sensitive enough to detect differences in early literacy comprehension behavior over time through the following research question: Are the comprehension constructs on the SOLR sensitive enough to capture change in behavior over time for students with ID and ASD?

Method

Participants

Participants included 12 students who took part in the FOTB intervention study for one academic year (Allor, Cheatham, & Al Otaiba., 2018; Allor et al., 2018). Participants were selected for the FOTB study according to the following criteria: (a) IQ between 40-79; (b) currently enrolled in grades 1 through 4; (c) verbal communication was primary means of communication; and (d) had limited literacy skills (Allor et al., 2018). The 12 participants include seven students with a primary diagnosis of ID or Down syndrome, three students with a primary diagnosis of ASD, and two students with a primary diagnosis of Smith Lemli-Opitz Syndrome, a developmental disability characterized by intellectual disability and/or learning and

behavior deficits. Participants range in age from 7-13 years with IQs ranging from 40-70 according to the *Wechsler Abbreviated Scale of Intelligence* (WASI-II; Wechsler, 2011). Additional demographic information, including *Peabody Picture Vocabulary Test-4* (PPVT-4; Dunn & Dunn, 2007) age equivalent scores, beginning and ending level of instruction in FOTB (i.e., level within the curriculum), and total dosage (intervention instructional time) is located in Table 1. All students who participated in the FOTB study demonstrated growth in word reading on a researcher-created proximal measure of taught sight words over the course of participation in the intervention (see Allor et al., 2018). Additionally, four of the students who progressed further into the intervention than the others were exposed to decodable words in addition to sight words, and these students also demonstrated growth on a proximal measure of decodable words (Allor et al., 2018).

Literacy Intervention

Video data was collected as students participated in the FOTB literacy intervention, which is designed for beginning readers and is inclusive of students with disabilities. The FOTB lessons are comprised of three major parts: 1) brief warm-up activities that teach phonemic awareness and word recognition skills, 2) story reading accompanied by dialogic reading questions, and 3) learning games that include comprehensive review of sounds, words, and skills taught in the intervention. The videos used for this study were taken during the book-reading portion of the lesson, which is designed to promote high levels of teacher/student interaction through turn taking during reading and opportunities to embed dialogic reading questions into instruction. Turn taking is supported through the design of the books, which include both student text, read by the student, and helper text, read by the teacher. The helper text supports meaning and complexity of the story, providing opportunities for discussion and abstract thinking. Picture

words are also included in the text to support meaning of the stories. These are words in the text with a small picture underneath the printed word; see Figure 1 for an example. Additionally, each teacher was provided with training in the FOTB intervention at the beginning of the school year, as well as bi-weekly coaching sessions with a trained research assistant (Allor et al., 2018).

Video Data Collection and Coding

Data for this manuscript was collected from videos of 12 students who participated in the FOTB intervention study for one academic school year. Data include one video from the beginning of the year and one from the end of the year, with about 8 months ($M = 8$, $SD = 4.8$) between video 1 and video 2 for each student. The video recordings capture the second portion of the intervention, story reading, which is designed to promote student-teacher interactions. Although lessons may be delivered to small groups, lessons in these videos were delivered to individual students. The story reading portion of the lesson lasted from about 5 to 14 minutes ($M = 9.16$, $SD = 4.07$); however, only the first 5 minutes of video was coded using the SOLR. Consistency of video length was important, as the coding procedures include 30-second time sampling, and videos of equal length could be similarly analyzed across students. Additional information about the video data is provided in Table 2.

Raters

Four trained graduate research assistants were responsible for coding all videos. The graduate research assistants were trained to use the SOLR by two members of the original SOLR development team, achieved at least 80% overall agreement with the gold standard established by the development team, and individually coded 8 of the 24 videos used in this analysis. Videos were assigned using a random number generator, and 33% of the videos were double coded for interrater reliability, which was calculated with the Gwet Agreement Coefficient (AC; Gwet,

2008). Average interrater reliability was $AC_1=0.83$, which is strong. Individual information for the 8 double coded videos (33%) is listed in Table 3.

Analysis

I used a paired samples t-test to compare the two sets of SOLR scores over time for each construct. Again, data are gathered from SOLR scores obtained from videos of students with disabilities during the story reading portion of the FOTB intervention from the beginning and end of their participation in the intervention. I hypothesized that there would be significant change in SOLR score over time and that some constructs would be more sensitive to change than other constructs.

Results

The research question asks if the comprehension constructs on the SOLR were sensitive enough to capture change in behavior over time for students with ID and ASD. Results from the t-test suggest that there was not a significant change in total SOLR score over time for any of the constructs. See Table 4 for results from the paired samples t-test, which examines change in SOLR score over time by construct. For four of the behavior categories (language development, elaborations, print, and off task/refusal) there was an increase in average score over time, although again, this change was not statistically significant. For three of the behavior categories (abstract thinking, engagement, and fluency/prosody) there was a decrease in average score over time, although again, this decrease was not statistically significant.

Examination of SOLR scores across individual participants shows that for four students there was a decrease in score across almost all behavior constructs over time, for three students there was an increase in score across behavior categories, and for five students the scores across time remained relatively consistent. CH is a clear example of a student for whom there was a

decrease in score across almost all behavior constructs from video 1 to video 2, and CF is an example of a student for whom there was an increase in score over time for each construct.

Possible explanations for the observed increase and decrease in behaviors over time are provided in the discussion. See Table 5 for SOLR score across individual behavior constructs at video one and video two, for each student.

Interrater reliability among raters was calculated for the 33% of videos that were double coded using the Gwet AC (Gwet, 2008). Across the eight double coded videos, the raters achieved an average of approximately $AC_1=0.83$. The strongest interrater reliability score was $AC_1=0.90$ (0.85; 0.95) for video 1 from KC. Again, information including Gwet AC estimate, standard error, and lower and upper confidence bound for each video that was double coded is found in Table 3.

Discussion

In this paper, I examined whether the comprehension constructs on the SOLR are sensitive enough to capture change in behavior over time for students with ID and ASD. According to the data that was gathered from students with disabilities participating in the FOTB intervention over the course of one academic year, the constructs on the SOLR were not sensitive to change over time. There was variability among change in score across the constructs, but this change in score was not significant. Further, only three of the twelve participants increased their SOLR score across almost every construct, while four student participants decreased across almost every construct, and five student participants demonstrated little change across constructs. In this discussion, I describe the videos used for SOLR analysis, which may not accurately depict the growth of the students who participated in the intervention. I purport that this is the most

likely explanation for the lack of significant changes in comprehension performance according to the SOLR.

The most likely explanation for the lack of significant change in SOLR construct over time is that the videos chosen for analysis are not an accurate representation of student comprehension growth in response to the FOTB intervention, which is in part due to the post hoc study design. The videos chosen for this analysis were collected during the study of FOTB, and at the time of data collection, the purpose of the videos was to document fidelity of teacher program implementation. When research assistants recorded the videos, the goal was not to capture reading comprehension behaviors but to monitor whether or not the teachers implemented all of the parts of the intervention (warm up, story reading, and learning activities) before the student moved on to the next level of the intervention. Activity during story reading changed depending on the day of the week the lesson was implemented. On some days, the teachers and students engaged in dialogic reading questions, creating opportunity for responses that exhibited language development, elaborations, or abstract thinking. However, on other days, students and teachers spent time making predictions at the beginning of the story reading, and then proceeded to spend the majority of their time together engaged in turn taking while reading from the book, which provided less opportunity for student responses. Video data was only recorded once per level for each student, without focus on previous exposure to the book or the frequency of teacher asked questions, as this was not the focus of the FOTB study. The present study was designed after the FOTB intervention was delivered, meaning all videos had already been collected, and due to the post hoc design, variability across video observations is likely to have affected the results.

Observations from Video Data

Upon examination of SOLR data at the individual student level (Table 5), it was clear that there was an increase in score across almost every construct over time for three students, a decrease in score across constructs for four students and relatively no change in score over time for five students. Interactions between the teachers and students, including the amount of teacher questioning and student response to book topic, which contributed to the amount of observed behaviors in each video, varied greatly from video 1 to video 2. This variation in observed behaviors affected the results of the t-test used to examine significant change in SOLR score over time.

Decreased Observed Behaviors over Time

Individual SOLR data from four students, including AM, CH, MS, and KR, indicated a decrease in observable behaviors over time, among almost every construct. For example, observed behaviors in most constructs decreased from video 1 to video 2 for CH. This was inconsistent with anecdotal evidence from research assistants, teachers, and parents that suggested growth in comprehension skills. CH is diagnosed with ID and has an IQ score of 43 according to the WASI-II (Wechsler, 2011). CH improved in sight and decodable word reading over the course of the intervention (see friendsontheblock.com for a graph of student data). Notable differences were observed in student/teacher interaction between video 1 and video 2 for CH that may have contributed to the decrease in SOLR score over time. In the first coded video, CH read a book from Level 1 of the intervention with his teacher. During this 5-minute video, the teacher provided many opportunities to engage in meaningful conversation about the book. His teacher began the story reading activity by asking two questions about the pictures, which immediately provided opportunities for him to talk about what he noticed on the page. Halfway through the five-minute video, the teacher had already asked CH six questions about the book

and the pictures. She asked both direct observation questions and open-ended questions that allowed the student to expand on his thinking, providing opportunity for elaborations and abstract thinking.

In the second video, CH was provided fewer opportunities to engage in conversation about the reading and the pictures than in video 1, which likely contributed to his decrease in score across almost every behavior construct. On the day that video 2 was filmed, the focus of the story reading lesson was on fluency, not dialogic reading questions. Across two of the behavior constructs, print and fluency, CH did show improvement over time, likely because in video 2 he spent a majority of the time reading directly from the page without questions from the teacher. During this video, CH began by reading directly from the book while his teacher sat next to him and provided corrective feedback when he struggled to read fluently. However, while CH read, she walked away to attend to other obligations in the classroom. During this time, CH continued reading fluently to himself, and when the teacher returned she praised him for staying on task, but did not ask CH any questions. She continued to monitor his reading and provided corrective feedback. Throughout the entirety of the video, she only asked one question about the book. She also provided less opportunity for elaborations, abstract thinking, and language development than was provided in video 1, as that day of the lesson focused on reading fluency and did not include dialogic questions embedded into the story reading.

Increased Observed Behaviors over Time

Individual SOLR data from several students, including CF, KC, and SM, indicated an increase in observable behaviors over time among almost every construct. For example, CF who is a student with Down syndrome and an IQ of 42 according to the WASI-II (Weschler, 2011),

made improvements in measures of sight and decodable word reading throughout participation in the intervention (see friendsontheblock.com for more information).

There were notable differences across the two coded videos which may have contributed to the increase in scores across construct for CF that differ from the observed behaviors in the videos for CH. In video 1, CF read from a book about making lunch. In the beginning of the video, she and the teacher talked about the picture words, but as an early reader, CF repeated the teacher for a majority of the conversation. During these exchanges, CF demonstrated language development, as she was talking out loud about the pictures; however, she did not demonstrate abstract thinking or elaborations, as she was not making connections or elaborating about what was on the page.

In video 2, CF had developed her word reading ability and was also much more enthusiastic about the book, which was about playing a game of hide-and-seek. Her interest was evident in the video, as she made many excited noises and comments, of which qualified as observed engagement behaviors. Additionally, she exhibited more fluency, as well as instances of language development, abstract thinking, elaborations, and print than she did in the first video. Additionally, her teacher asked more dialogic reading questions that provided opportunity for conversation than in the first video, which may explain the increase in observed behaviors in the abstract thinking and elaborations constructs.

There were notable differences between the videos that appear to have influenced the amount of observed behaviors that were captured for CF over time, many of which were due to the amount of dialogic reading questions asked by her teacher in video 2, increasing her opportunities for response. This increase in observed behavior was not captured in the video footage for CH, even though comprehension behavior growth was observed in-person for the

student over the course of his participation in the intervention. The reason the behavior growth was not evident was likely due to the activity that was embedded into the lesson on the day that the footage was captured for video 2, as the video captured a day of the week in which dialogic reading questions were not the focus of story reading.

Limitations and Directions for Future Research

Many of the limitations associated with this study could be corrected with careful consideration to the collection of the videos in future research on the SOLR. To accurately assess whether or not the SOLR is sensitive to change, future research should be designed in which researchers create a detailed plan for collecting video data that minimizes variables that have the potential to affect opportunities to respond so that opportunities are equal across videos. Additionally, included in the plan for data collection should be information regarding more than two videos per student (beginning and end of intervention). When monitoring the progress of students with ID and ASD there is often great variation among the scores of even one student, and more than two data points may be necessary to show the most accurate trend in growth (Allor et al., 2013; Jones et al., 2018; Wallace, Ticha, & Gustafson, 2010).

Additionally, among the several teachers who implemented the FOTB instruction across the videos, there were varying degrees of experience, and they also implemented the FOTB instruction with different levels of fidelity. Some teachers are more purposeful about asking dialogic reading questions and incorporating questions with different degrees of difficulty into the student-teacher interactions, providing opportunity for higher-level thinking, even on the days of the lesson where dialogic reading questions are not embedded into the activity. In future research exploring the sensitivity of the SOLR, teachers should be selected who have the same level of experience or training in the intervention used in the study and be provided with training

on the behaviors listed on the SOLR; this would rule out any differences due to implementation of curriculum.

In the future, an adapted version of the SOLR could be created for use during the school year to intensify comprehension instruction and target one comprehension construct at a time. Ideally, if the tool were used in conjunction with reading comprehension instruction, the teacher(s) would be able to target specific behaviors or comprehension constructs in which the student is weak to address growth in that specific area. An intervention could be implemented for which the target behavior is one of the seven comprehension constructs listed on the SOLR, meeting mastery of one construct at a time. In order to target one construct at a time, a shorter, more feasible version of the SOLR could be developed and used as a progress monitoring tool. Possibilities for this shorter version of the SOLR include a checklist or Likert scale instead of a 30-second interval system, or a version that only focuses on one or two of the specific constructs at a time, as opposed to all seven.

As the SOLR is inclusive of early/developing readers of all abilities, future lines of inquiry could focus on specific populations, such as struggling readers, students with learning disabilities, dyslexia, or English language learners. A focus on specific groups of students could potentially influence the development of successful interventions or instructional techniques that support groups of struggling readers with differing needs.

Potential Impact of the Study

The goal of this study was to examine whether or not the SOLR is sensitive enough to capture change in comprehension behaviors for students with disabilities, as there is a need for comprehension measures that are inclusive of this population (see Conner, 2020a). A team of researchers created the SOLR, using the Systematic Assessment of Book Reading (Pentimonti et

al., 2012) as a guide, with the goal of developing a tool that could capture the comprehension growth that was observed by researchers and teachers throughout the FOTB intervention (Conner, 2020b). Using data from FOTB study (Allor et al., 2018; Allor et al., 2020), the goal of this study was to examine the already collected video data from a new perspective using a reading comprehension measure that would potentially be sensitive enough to capture change in comprehension behaviors for the student participants.

Currently, most measures of reading comprehension behavior that contain reading passages paired with inferential or literal recall questions do not address other types of comprehension behavior that are necessary to develop reading comprehension such as engagement, oral language or listening comprehension (Yuill & Oakhill, 1991, Foorman et al., 2018). Development of the SOLR could potentially influence the field of special education research by providing an accurate measure of reading comprehension behaviors that is appropriate for beginning readers and inclusive of students with intensive needs, such as ID and ASD. In the examination of technical adequacy of the SOLR, Conner (2020b) details the development of the SOLR, including the strong interrater reliability and construct validity of the tool. Based on examination of the technical adequacy of the tool, the SOLR may be more sensitive to change than was evident in this study (Conner, 2020b). Further research is needed that addresses the limitations described.

The possibility for a progress monitoring tool created from an adapted version of the SOLR is another important path of future inquiry. Currently, there is a need for more research on the development of progress monitoring tools that measure reading comprehension behaviors, as accurate comprehension progress monitoring tools are critical to improve both research of comprehension programs and teacher delivery of reading comprehension instruction to students

with disabilities. Educators and researchers alike would benefit from reliable and valid comprehension measures to assess the skills that are targeted during intervention development as well as to assess responsiveness for students with disabilities. The development of the SOLR, described in this paper, is likely to address gaps in the field of research as it has the potential to measure the cascading levels of reading comprehension behaviors with the ultimate goal of supporting growth for all early readers.

References

- Allor, J. H., Cheatham, J., & Al Otaiba., (2018). *Friends on the block*. Dallas, TX: Friends on the Block.
- Allor, J. H., Gifford, D. B., Al Otaiba, S., Miller, S. J., & Cheatham, J. P. (2013). Teaching students with intellectual disability to integrate reading skills: Effects of text and text based lessons. *Remedial and Special Education, 34*, 346–356.
doi:10.1177/0741932513494020
- Allor, J. H., Mathes, P. G., Roberts, J. K., Cheatham, J. P., & Otaiba, S. A. (2014). Is scientifically based reading instruction effective for students with below-average IQs?. *Exceptional Children, 80*, 287-306.
- Allor, J., Yovanoff, P., Al Otaiba, S., Ortiz, M., & Conner, C. (2020). Evidence for a literacy intervention for students with intellectual and developmental disabilities. *Education and Training in Autism and Developmental Disabilities, 55*, 290-302.
- Browder, D. M., Ahlgrim-Delzell, L., Courtade, O., Gibbs, S. L., & Flowers, C., (2008). Evaluation and the effectiveness of an early literacy program for students with significant developmental disabilities. *Exceptional Children, 75*, 33-52.
- Conner, C. (2020a). A synthesis of reading comprehension measures for students with autism spectrum disorder and low IQ. (Unpublished doctoral dissertation). Southern Methodist University, Texas.
- Conner, C. (2020b). Examining the technical adequacy of the systematic observation of language and reading. (Unpublished doctoral dissertation). Southern Methodist University, Texas.
- Connors, F. A. (1992). Reading instruction for students with moderate mental retardation: Review and analysis of research. *American Journal of Mental Retardation, 96*, 577-597.

- Connor, C. M., Alberto, P. A., Compton, D. L., & O'Connor, R. E. (2014). Improving Reading Outcomes for Students with or at Risk for Reading Disabilities: What Have We Learned? *Synthesis of IES Research*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9, 561–584.
- Dunn, L. M., Dunn, D. M. (2007) *PPVT-4: Peabody picture vocabulary test* Minneapolis, MN. Pearson Assessments.
- El Zein, F., Solis, M., Lang, R., & Kim, M. K. (2016). Embedding perseverative interest of a child with autism in text may result in improved reading comprehension: A pilot study. *Developmental Neurorehabilitation*, 19, 141-145.
doi:10.3109/17518423.2014.915893ESSA, 2015
- Foorman, B. R., Arndt, E. J., & Crawford, E. C. (2011). Important constructs in literacy learning across disciplines. *Topics in Language Disorders*, 31, 73-83.
doi:10.1097/TLD.0b013e31820a0b86
- Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 1, 29-48.
- Jones, F. G., Gifford, D., Yovanoff, P., Al Otaiba, S., Levy, D., & Allor, J. (2018). Alternate assessment formats for progress monitoring students with intellectual disabilities and

- below average IQ: An exploratory study. *Focus on Autism and Other Developmental Disabilities*, 34, 41-51. doi:1177/1088357618762749
- Katims, D. S. (2001). Literacy assessment of students with mental retardation: An exploratory investigation. *Education and Training in Mental Retardation and Developmental Disabilities*, 363-372.
- Ketterlin-Geller, L. R. (2008). Testing students with special needs: A model for understanding the interaction between assessment and student characteristics in a universally designed environment. *Educational Measurement: Issues and Practices*, 27, 3–16
- Knight, V. F., Wood, C. L., Spooner, F., Browder, D. M., & O'Brien, C. P. (2015). An exploratory study using science eTexts with students with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities*, 30, 86-99. doi:10.1177/1088357614559214
- Knott, Patterson & Hodges, 1997;
- Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology*, 95, 211. doi:10.1037/0022-0663.95.2.211
- Lemons, C. J., & Fuchs, D. (2010). Modeling response to reading intervention in children with Down syndrome: An examination of predictors of differential growth. *Reading Research Quarterly*, 45, 134–168.
- Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*, 73, 139-158. doi: 10.1006/jecp.1999.2498
- Pentimonti, J. M., Zucker, T. A., Justice, L. M., Petscher, Y., Piasta, S. B., & Kaderavek, J. N. (2012). A standardized tool for assessing the quality of classroom-based shared reading:

- Systematic Assessment of Book Reading (SABR). *Early Childhood Research Quarterly*, 27, 512-528.
- Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. *Reading and Writing*, 4, 245–256.
- Tarlow, K. R. (2016). An improved rank correlation effect size statistic for single-case designs: Baseline corrected Tau. *Behavior Modification*, 41, 427-467.
- Turner, H., Remington, A., & Hill, V. (2017). Developing an intervention to improve reading comprehension for children and young people with autism spectrum disorders. *Educational and Child Psychology*, 34, 13-26.
- Wallace, T., Tichá, R., & Gustafson, K. (2010). Technical characteristics of general outcome measures (GOMs) in reading for students with significant cognitive disabilities. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 26, 333–360
- Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II)*. San Antonio, TX: NCS Pearson.
- Yuill, N., & Oakhill, J. V. (1991). *Children's problems in text comprehension*. Cambridge, UK: Cambridge Univ. Press.

Table 1*Participant Demographic Information*

Case	Age	IQ	Starting Level	Ending Level	PPVT	Disability	Total
					Age Eq		Dosage (hours)
AMM	9	42	6	9	2:06	ASD	26.48
CF	9	42	1	3	2:09	Down Syndrome	30.46
CH	10	43	1	7	5:04	ID/SI	57.45
CS	7	55	1	3	3:04	ASD/SI	20.50
EM	6	59	1	4	2:04	Smith-Lemli Opitz	31.96
JG	10	56	1	3	3:04	Down Syndrome	50.15
KC	10	47	6	7	3:06	Down Syndrome	57.08
KR	9	41	3	5	2:07	ID/SI	38.50
LB	9	70	1	3	5:03	ASD/ID/SI	32.10
LD	12	65	8	14	5:04	ID/SI	40.58
MS	10	65	7	9	5:04	ID/ED	31.20
SM	10	40	1	6	2:03	Smith-Lemli Opitz	30.66

Table 2*Video Data Information*

Student	Level	Video 1 date	Video 1 length	Level	Video 2 date	Video 2 length
AMM	4	11/14/14	5:00	9	3/16/16	5:40
CF	1	11/4/14	5:50	5	4/20/15	7:30
CH	1	03/03/16	6:21	7	2/13/17	20:26
CS	1	11/14/16	6:50	3	3/9/17	5:00
EM	1	10/8/14	6:20	4	3/29/16	9:58
JG	1	10/8/14	14:00	4	5/6/15	6:00
KC	6B	11/6/14	6:52	8B	5/7/15	7:00
KR	3	10/13/16	10:00	5	3/7/17	14:20
LB	1	11/11/16	5:50	3	3/24/17	6:45
LD	9	4/19/16	12:25	13	12/15/16	5:00
MS	7	3/3/16	15:34	9	5/26/16	12:58
SM	1	11/4/14	8:30	6B	10/14/15	10:25

Table 3

Interrater reliability information from video coding

[illegible]

Table 4*Paired Samples T-Test*

Paired Differences								
Behavior Category	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower 95% Confidence Bound	Upper 95% Confidence Bound			
Language Development	2.16	7.59	2.19	-2.66	6.99	0.99	11	.344
Abstract Thinking	-0.67	1.61	0.47	-1.69	0.36	-1.43	11	.180
Elaborations	0.91	3.70	1.07	-1.44	3.27	0.86	11	.410
Print	0.33	3.75	1.08	-2.05	2.72	0.31	11	.764
Engagement	-1.25	4.48	1.29	-4.09	1.59	-0.97	11	.354
Fluency/Prosody	-3.00	7.65	2.21	-7.86	1.86	-1.36	11	.202
Off task/refusal	0.25	1.22	0.35	-0.52	1.02	0.71	11	.491

Table 5*Individual SOLR scores: video 1 and video 2*

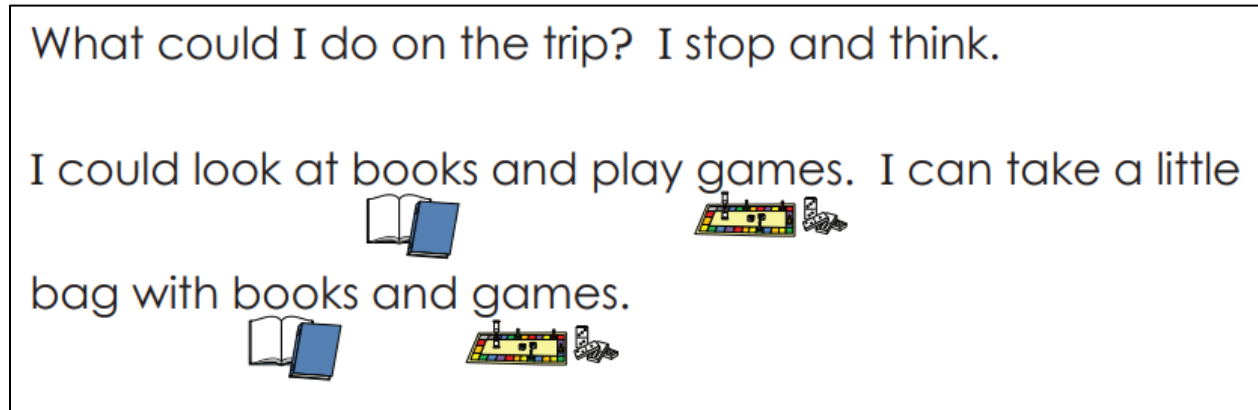
Participant	Count of Observed Behaviors by Construct						
	Language Development	Abstract Thinking	Elaborations	Print	Engagement	Fluency/Prosody	Off Task/refusal
AM							
video 1	8	0	0	20	21	23	0
video 2	4	0	0	16	15	16	1
CF							
video 1	5	0	4	16	15	12	0
video 2	11	2	5	21	22	20	1
CH							
video 1	15	1	0	18	24	23	0
video 2	0	0	0	20	20	28	0
CS							
video 1	7	0	4	15	19	14	3
video 2	2	1	1	17	18	16	2
EM							
video 1	13	0	4	19	25	26	0
video 2	12	0	1	14	25	18	0
JG							
video 1	8	0	2	18	20	16	0
video 2	16	2	4	16	23	15	0
KC							
video 1	3	0	0	16	15	14	0
video 2	10	1	4	12	22	24	0
KR							
video 1	16	0	0	20	23	18	3
video 2	8	1	1	14	24	10	0
LB							

Count of Observed Behaviors by Construct

Participant	Language Development	Abstract Thinking	Elaborations	Print	Engagement	Fluency/Prosody	Off Task/refusal
video 1	12	4	7	16	22	12	0
video 2	14	1	9	16	21	16	1
LD							
video 1	18	0	5	20	22	23	0
video 2	4	0	1	23	20	35	0
MS							
video 1	16	0	12	12	19	13	0
video 2	13	2	2	13	22	18	0
SM							
video 1	11	0	4	15	16	6	2
video 2	12	3	3	19	24	20	0

Figure 1

Image of picture words from *Friends on the Block* (Allor et al., 2016): Level 7 book “Going on a trip”



Chapter V: Conclusion

The purpose of this three-paper dissertation is to describe the need for the Systematic Observation of Language and Reading (SOLR), a reading comprehension measure designed to be inclusive of students with Intellectual Disability (ID) and Autism Spectrum Disorder (ASD). Research supports students with ID and ASD responding positively to comprehensive reading instruction; however, further research on reading comprehension, including effective measures, are needed for this population (Allor et al., 2014; El Zein et al., 2016; Turner et al., 2017). Many assessments of comprehension are narrow, with limited response formats, and do not capture other aspects of reading development such as engagement, oral language, and listening comprehension that are necessary for reading comprehension (Fletcher et al., 2016; Francis et al., 2006).

The first paper in this dissertation, a systematic review, examines the comprehension measures used in experimental studies to assess comprehension growth in students with ASD and low IQ. The research questions asked what types of measures were used to assess growth in the studies and what types of comprehension constructs these measures assessed. Of the studies that met criteria for the synthesis, most measures included were designed by the researchers to assess growth through question probes asking literal recall or inferential questions. Additionally, five different constructs of comprehension were assessed through the multiple measures used across studies (see Conner, 2020a). In the discussion of paper 1, the need for a comprehension measure that assesses oral language, engagement, and other behaviors necessary for reading comprehension and is sensitive enough to capture growth in students with ID and ASD is described.

The second paper in this dissertation describes the technical adequacy of the SOLR, a reading comprehension measure that includes seven constructs of reading comprehension and is inclusive of students with ID and ASD. The SOLR includes 25 behaviors that are each aligned to one of the seven constructs. Each of the behaviors are either observed or not observed at 30-second intervals through video-recorded literacy instruction of students with ID and ASD. To examine the technical adequacy of the SOLR, interrater reliability and construct validity were assessed. Results support strong overall interrater reliability, which varied by construct; however, within that variation, interrater reliability for each construct remained relatively strong. Results from the survey designed to assess construct validity also support strong overall validity, which varied by behavior and construct (see Conner, 2020b). In the discussion section of this paper, future research with the SOLR, including examining the sensitivity of the tool across students with ID and ASD in response to reading interventions, is described.

The third paper in this dissertation examines the sensitivity of the SOLR over time, assessing growth of a group of students with ID and ASD over the course of participation in a comprehensive literacy intervention designed for students with disabilities. To explore sensitivity, video data of 12 students with disabilities were coded using the SOLR, including one video from the beginning of instruction and one video from the end of instruction. Paired samples t-tests were used to examine significant change in constructs over time across the students; however, results found no significant change in any construct (see Conner, 2020c). In the discussion section of this paper, video data for two specific students are described; one student who showed improvement in each construct over time and one student who showed regression in most constructs over time. Description and analysis of the interactions in the videos suggests the reason for the lack of significant findings is likely due to differences in video data

from video 1 to video 2 for each student. As this study was designed post hoc, the data was collected from videos that were not taken for the purpose of capturing comprehension behaviors. Therefore, there is great variation in the lesson delivery that is captured in each video. During some lessons, teachers asked a variety of dialogic reading questions, which created opportunities for most of the behaviors listed on the SOLR to be observed. During other lessons, teachers were more focused on the student reading fluently and finishing the story, in which opportunities for many of the behaviors listed on the SOLR were limited. The need to design a study in which video data are collected for the purposes of measuring behavior is described in the implications for future research.

This dissertation has the potential to affect the field of reading comprehension research by providing a sensitive measure designed to capture growth of students with intensive needs, such as ID and ASD. The need for the SOLR is described in the synthesis included in this dissertation (Conner, 2020a). Additionally, implications for future research on the SOLR, both to refine the tool so that it becomes more technically sound (Conner, 2020b) and to examine the sensitivity with a study design that includes purposeful data collection (Conner, 2020c), are described in this three-paper dissertation.

References

- Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Allor, J. H., Mathes, P. G., Roberts, J. K., Cheatham, J. P., & Otaiba, S. A. (2014). Is scientifically based reading instruction effective for students with below-average IQs? *Exceptional Children*, 80, 287-306. doi: 10.1177/0014402914522208
- Bursuck, W. D., & Darnier, M. (2011). *Reading instruction for students who are at risk or have disabilities* (2nd ed.). Boston, MA: Pearson Education, Inc.
- Chiang, H., & Lin, Y. (2007). Reading comprehension instruction for students with autism spectrum disorders: A review of the literature. *Focus on Autism and Other Developmental Disabilities*, 22, 259-267.
- Conner, C., (2020a). A synthesis of reading comprehension measures for students with autism spectrum disorder and low IQ (Unpublished doctoral dissertation). Southern Methodist University, Texas.
- Conner, C., (2020b) Examining the technical adequacy of the systematic observation of language and reading (Unpublished doctoral dissertation). Southern Methodist University, Texas.
- Conner, C., (2020c) Exploring the Sensitivity of the Systematic Observation of Language and Reading across Students with intellectual disability and autism spectrum disorder (Unpublished doctoral dissertation). Southern Methodist University, Texas.
- Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10, 277-299.

- El Zein, F., Solis, M., Lang, R., & Kim, M. K. (2016). Embedding perseverative interest of a child with autism in text may result in improved reading comprehension: A pilot study. *Developmental Neurorehabilitation, 19*, 141-145.
doi:10.3109/17518423.2014.915893
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10*, 323-330.
- Foorman, B. R., Arndt, E. J., & Crawford, E. C. (2011). Important constructs in literacy learning across disciplines. *Topics in Language Disorders, 31*, 73-83.
doi:10.1097/TLD.0b013e31820a0b86
- Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading, 10*, 301-322.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.
- Mastropieri, M. A., & Scruggs, T. E. (1997). Best practices in promoting reading comprehension in students with learning disabilities 1976 to 1996. *Remedial and Special Education, 18*, 198-213. doi:10.1177/074193259701800402
- Mims, P. J., Hudson, M. E., & Browder, D. M. (2012). Using read-alouds of grade-level biographies and systematic prompting to promote comprehension for students with moderate and severe developmental disabilities. *Focus on Autism and Other Developmental Disabilities, 27*, 67-80. doi:10.1177/1088357612446859
- Perfetti, C. A. (1985). Reading ability. New York: Oxford Press

Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development*, 61, 1728–1743.

Turner, H., Remington, A., & Hill, V. (2017). Developing an intervention to improve reading comprehension for children and young people with autism spectrum disorders. *Educational and Child Psychology*, 34, 13-26.

Appendix A

Image of the Systematic Observation of Language and Reading (SOLR) codebook

Construct	Student Coding categories	Definition
Language Development	Responds using complete meaningful thought	The student initiates a conversation OR responds to a question either using a complete thought or a complete sentence. A complete thought may be "Red hat; Mom sick in bed; No like it." etc.
	Labels or describes story actions	The student names, refers to, or describes something that is happening on the current page of the story. This specifically refers to an action (a verb). This can be student initiated or prompted by a question that the teacher/tutor asks. Please note in the comments section whether it was prompted by the student or by the adult.
	Labels or describes nouns, locations, characters	The student names, refers to, or describes a person, place, or character on the current page of the story. This specifically refers to a noun. This can be student initiated or prompted by a question that the teacher/tutor asks. Please note in the comments section whether it was prompted by the student or by the adult.
	Provides description/characteristics	The student names, refers to, or describes something other than a verb or a noun on the current page of the story. This can be a reference to a characteristic, a description, and emotion or other adjective. This can be student initiated or prompted by a question that the teacher/tutor asks. Please note in the comments section whether it was prompted by the student or by the adult.
	Gives a definition for a word	The student provides the definition to a word. This can be either student initiated or prompted by an adult. Please note in the comments section whether it was prompted by the student or by the adult.
Abstract Thinking	Is able to compare or contrast	The student either compares or contrasts two or more things. This may be an explicit instance of providing similarities or differences or can also be a mention of something that the student is comparing to what is currently being read (for this the teacher will usually provide clarity or reinforcement on the idea).

	Offers an understanding from the character's point of view	The student demonstrates that he/she understands the character's point of view. This could be student initiated or prompted by the teacher. This could also be a comparison that the student provides to his/her own life that is similar to the author's point of view.
	Makes a prediction or hypothesis about future events	The student makes a prediction or a hypothesis about something that may happen later in the book/story. This may be student initiated or teacher prompted.
	Offers an inference, analysis, or explanation	The student offers/provides an explanation of the event in the book not about a specific word. They may make an inference about a character or something that is happening. This may be student initiated or teacher prompted. Please note in the comments section whether it was prompted by the student or by the adult.
Elaboration	Makes a text to life connection	The student makes a comment or response that provides a connection between the current events on the page and a life event for the student.
	Describes/ recalls information about text from prior pages or from previous readings	The student makes a connection between what is going on on the page to something else that they have read within the SAME book from current or an earlier reading (i.e., earlier that day or from a previous reading of the book). This may have to be confirmed by an adult in order to know for sure that they are mentioning prior text or from a previous reading.
	Extends conversation with relevant question or story remarks	The student elaborates on a question or initiates a story with or asks a question based on what he/she sees on the current page. The purpose of this code is to see whether the student provides an extension rather than just simply answering a question at hand.
	Makes a text to text connection between books read	The student makes a connection between what is going on on the page to circumstances or characters that they have read about in a DIFFERENT book. This may have to be confirmed by an adult in order to know for sure that they are mentioning a previous page or reading.
Print	Is involved in turn taking with text reading- verbalizing	The student does not read while the teacher is reading and then reads or tries to read when it is his/her turn.
	Demonstrates understanding of one-to-one correspondence.	The student points to a word while reading or the teacher points and student reads the word/letter correctly. The student shows that he/she understands
	Initiates reading without prompting	The student starts reading without any prompting by the teacher. Prompting can be verbal or can be physical such as the teacher pointing to the word. If there is a wait then a prompt the opportunity was there but not taken. If the teacher just automatically prompts without waiting then there is no opportunity.

Engagement	Demonstrates excitement/engagement with the illustrations/pictures	The student shows engagement with the pictures. This can be physical such as looking at the pictures intently or pointing or can be verbal such as making a comment or asking a question about the pictures/illustration. This can occur before the book reading even starts. This code is all about the pictures and NOT about the text or storyline.
	Demonstrates excitement/engagement with story, character or actions in the book	The student shows engagement with the text or story line. This can be physical such as looking at the words or listening intently or can be verbal such as making a comment or asking a question about the story. This is about the story or what is happening in the book or with the characters.
	Physically engages with the book by holding it, pointing within it, and visually attending to it	The student is physically involved with the book. This may look like the student holding the book, pointing to the book, visually attending to the book, turning the pages, etc. In order to be coded this needs to occur for 15 seconds or more.
Fluency/ Prosody	Reads in meaningful phrases	The student reads the words on the page with automaticity or in smooth connected sections. Notably, the student reads the words smoothly, with a rhythm, and without distinct pauses between each word. Prompting is unnecessary.
	Reads words accurately or with minimal errors and/or prompting	The student reads the words on the page in a fluent manner, using accurate decoding with minimal errors. Please note in the comments section whether the adult prompted student to say a word.
	Student independently attempts to self-correct, repeat to problem solve, and/or sound out words	The student attempts to go back or correct him/her-self when making an error during text reading. This is to be initiated by the student and not by the teacher.
	Uses pitch, stress, and intonation to convey meaning	The student clearly is reading with varying pitch, stress, or intonation on certain words to convey meaning in the story.
Off Task/ Refusal	Refuses to participate in reading (e.g., verbalizes refusal, puts head on desk, tries to escape)	The student refuses to participate in the activity/reading. This may look like the student saying no or I don't want to. This may also be the student putting his head on the desk looking away from the story, hiding under the table, trying to run away, etc. In order to be coded this needs to occur for 15 seconds or more.
	Visually disengages with text or task by looking away from teacher or book	The student looks away from the text, closes his/her eyes, is looking somewhere else in the room. In order to be coded this needs to occur for 15 seconds or more.

Image of the Systematic Observation of Language and Reading (SOLR) scoring sheet

Book title:		Student:										Group:																						
Teacher/ Paraprofessional		Date:										Begin time:										End time:												
Constructs	Behaviors/Items	0-30	1-00	1-30	2-00	2-30	3-00	3-30	4-00	4-30	5-00	5-30	6-00	6-30	7-00	7-30	8-00	8-30	9-00	9-30	10-00	10-30	11-00	11-30	12-00	12-30	13-00	13-30	14-00	14-30	###	###	###	Comments
Language Development	Responds using complete meaningful thought																																	
	Labels or describes story actions																																	
	Labels or describes nouns, locations, characters																																	
	Provides description/characteristics																																	
	Gives a definition for a word																																	
Abstract Thinking	Is able to compare or contrast Offers an understanding from the character's point of view																																	
	Makes a prediction or hypothesis about future events																																	
	Offers an inference, analysis, or explanation																																	
Elaborations	Makes a text to life connection Describes/ recalls information about text from prior pages or previous readings																																	
	Extends conversation with relevant question or story remarks																																	
	Makes a text to text connection between books read																																	
Print	Is involved in turn taking with text reading- verbalizing																																	
	Demonstrates understanding of one-to-one correspondence.																																	
	Initiates reading without prompting																																	
Engagement	Demonstrates excitement/engagement about reading the illustrations/pictures																																	
	Demonstrates excitement/engagment about story, character or actions in the book																																	
	Physically engages with the book by holding it, pointing within it, and visually attending to it																																	
Fluency, Prosody	Reads in meaningful phrases																																	
	Reads words accurately or with minimal errors and/or prompting																																	
	Student independantly attempts to self-correct, repeat to problem solve, and/or sound out words																																	
Off Task/ Refusal	Uses pitch, stress, and intonation to convey meaning																																	
	Refuses to participates in reading (e.g., verbalizes refusal, puts head on desk, tries to escape)																																	
	Visually disengages with text or task by looking away from teacher or book																																	

Appendix C

Image of qualtrics survey used for internal validity

Page 1: Informed consent

SMU EXEMPT LEVEL PARTICIPATION EXPLANATION AND CONSENT SCRIPT

I am conducting a research study to learn more about an observation tool to assess reading comprehension for students with disabilities called the Systematic Observation of Language and Reading (SOLR). Your participation in this study is voluntary. If you agree to take part and then change your mind, you can withdraw for any reason. There are no penalties if you withdraw, decline to participate, or skip any parts of the study (survey). If you agree to participate, you will be asked to complete an online survey. Your participation should take about 10 minutes. There are no risks associated with participation in this research and no identifying data will be collected. After all data has been collected, you will be provided with a link to a completed version of the assessment tool along with a description of how to use the assessment to inform comprehension instruction in the classroom.

If you would like to participate in this research study, please enter your initials in the space below:



Image of qualtrics survey used for internal validity

Page 2: Behavior and construct matching (image 1)

Match each item listed on the left to the correct comprehension construct in the box on the right by dragging the item to the appropriate box. Note: There will not be an even number of items in each box!

Items

Is able to compare or contrast

Makes a prediction or hypothesis about future events

Makes a text to life connection

Labels or describes nouns, locations, characters

Describes/recalls information about text from prior pages or previous readings

Initiates reading without prompting

Gives a definition for a word

Language Development - the ability to express language**Abstract Thinking - higher level cognitive processes**

Page 2: Behavior and construct matching (image 2)

Offers an understanding
from the character's point
of view

Demonstrates
excitement/engagement
about reading the
illustrations/pictures

Uses a complete
meaningful thought

Makes a text to text
connection between
books read

Offers an inference,
analysis, or explanation

Extends conversation with
relevant question or story
remarks

Physically engages with
the book by holding it,
pointing within it, and
visually attending to it

**Elaborations - ability to make
connections between what the
student knows and what is present
in the book**

**Print - understanding that print has
form and function**

Page 2: Behavior and construct matching (image 3)

Demonstrates
understanding of one-to-
one correspondence

Labels or describes story
actions

Demonstrates
excitement/engagement
about story, character, or
actions in the book

Provides
description/characteristics

Visually disengages with
text or task by looking
away from teacher or
book

Reads in meaningful
phrases

Reads words accurately
or with minimal errors
and/or prompting

**Engagement - attending to the book
and pictures**

**Fluency/Prosody - reading with
automaticity, accuracy, and
expression**

Page 2: Behavior and construct matching (image 4)

Refusal to participate in reading (e.g., verbalizes refusal, puts head on desk, tries to escape)

Is involved in turn taking with text reading-verbalizing

Student independently attempts to self-correct, repeat to problem solve, and/or sound out words

Uses pitch, stress, and intonation to convey meaning

Off Task/Refusal - student exhibits behaviors that interfere with ability to learn

