The reliability and validity of a progress-monitoring tool: A psychometric examination of the phonological awareness skills of preschoolers with ASD

Jay R. Martini

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

James Mazza, Chair

Roxanne Hudson

Raphael Bernier

Program Authorized to Offer Degree:

College of Education: Educational Psychology - School Psychology

University of Washington

Abstract

The reliability and validity of a progress-monitoring tool: A psychometric examination of the phonological awareness skills of preschoolers with ASD

Jay R. Martini

Chair of the Supervisory Committee:
Dr. James Mazza
Educational Psychology

The purpose of this study was to conduct a psychometric evaluation the *Sound Beginning* phonological awareness progress monitoring tool. This assessment was used to track emergent literacy skills of preschoolers with autism spectrum disorder who were participating in a randomized trial studying early literacy interventions. Research questions in the present study included an investigation of levels of reliability, validity, item level data and the measure's sensitivity to change in phonological awareness skills. The present study included a sample of 40 individuals with autism spectrum disorder who were in their last year of preschool. The progress monitoring tool was administered up to four times during intervention. Furthermore, a variety of emergent literacy and oral language measures were administered before and after intervention and were used in this analysis. The results of this study suggest that the phonological awareness progress monitoring assessment demonstrated sufficient levels of reliability, validity, and sensitivity to change when used with this sample of preschoolers with autism spectrum disorder.

Keywords: Phonological Awareness, Emergent Literacy, Psychometrics, Validity, Reliability, Autism Spectrum Disorder

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DEDICATION

This dissertation is dedicated to my loving family and supportive partner. To my family, who encouraged my curiosity from a young age and have supported me over the years. To my mother, who was an integral team member through the dissertation process. Finally to my partner, whose support was unending during graduate school.

Chapter I: Introduction

In the field of education, when teaching academic skills, valid and reliable measures are important for monitoring learning. Measurement is a broad process for quantifying behavior, whereas testing is a specific and more formal type of measurement (Allen & Yen, 2002). Testing is considered to be a systematic procedure for collecting samples of behavior that have been identified by an operational definition within a particular construct domain (Crocker & Algina, 1986). Within the field of educational and psychological assessment, the commonly accepted procedures used for the examination of testing measures were set forth by a joint committee of the American Education Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement Education (NCME) (2014).

Learning to read is an important and complex skill that people must learn to function in our society (Adams, 1990; Whitehurst & Lonigan, 2001). Emergent literacy sets the foundation for later reading and relies upon the individual child's experiences, precursor abilities, knowledge and aptitudes (Whitehurst & Lonigan, 1998). Children with Autism Spectrum Disorder (ASD) present with a wide range of reading abilities and include many individuals who will likely need extra support to attain future academic success (Nation et al., 2006). While the majority of children with ASD have deficits in reading comprehension with intact decoding skills, a sizeable portion demonstrate challenges with both decoding and comprehension (Davidson & Weismer, 2014; Nation et al., 2006). Many children with ASD will primarily require supports to improve reading comprehension, but most may also benefit from phonological awareness training to improve later decoding skills.

Scholars have used various approaches to describe reading and to categorize children's typical and atypical reading development, and one such model is The Simple View of Reading

(Gough & Tunmer, 1986). This empirically based model is useful because it provides a straightforward framework to classify which processes may be getting in the way of reading development. The Simple View allows a teacher to divide reading into two parts; decoding, which is a skill unique to reading, and comprehension. Phonological awareness, an emergent literacy skill, has been demonstrated to have predictive correlation with future reading success and maps closely with decoding skills in typically developing children. (National Early Literacy Panel, 2008; Wagner et al., 1997). While reading success varies by age and within schools, some children who experience developmental differences or mental health disorders may find reading and learning to read quite challenging, especially those who are children with Autism Spectrum Disorders (ASD).

The purpose of this study was to evaluate the reliability and validity of the use of the *Sound Beginning* Phonological Awareness Assessment (SBPAA) measure (Lane & Pullen, 2004) when used to track the progress and growth of Phonological Awareness (PA) in children with ASD. The validation study examined a dataset that was produced as part of a larger literacy study of prekindergarten children with ASD. Kane's framework was used as guide to assess levels of validity (Kane, 2004, 2006, 2013). Many studies regarding the validation of literacy assessments do not examine a measure's validity for use with specific subpopulations. Today, one in every 68 children is expected to be diagnosed with ASD (Christensen et al., 2016). Researchers, led by Nation (2006), have shown that children with ASD are a heterogeneous population in regards to academic and reading ability. Various types of challenges with reading are not uncommon in this population of children, and it is important to accurately assess these reading difficulties in order to inform appropriate treatment and intervention (Davidson & Ellis Weismer, 2014; Gabig, 2010; Nation et al., 2006). While emerging research suggests that oral

language is the driving force for reading deficits, specifically reading comprehension difficulties, in children with ASD, some research has shown that deficits in phonological awareness are a contributing factor and interventionists may be targeting the wrong component of reading in this population (Davidson & Ellis Weismer, 2014). Thus, valid instruments are essential to assess and track phonological awareness skills of children with ASD, as well for all children.

This study is an investigation of the levels of reliability and validity of the SBPAA when used with preschoolers with ASD. Along with this over-arching investigation, secondary analyses were conducted to examine item level performance and explore the assessment's ability to detect change. Detailed explanation of the questions, the rationale behind the questions, and their respective analyses are discussed.

Chapter II: Literature Review

Federal initiatives have called for increased accountability in education through the No Child Left Behind Act (NCLB; 2001), the reauthorization of the Individuals with Disabilities Education Improvement Act (IDEA; 2004), and now the Every Student Succeeds Act (ESSA; 2015) which replaced NCLB. The language in IDEA emphasizes the need for sound interventions and instruction that are based on scientifically defensible research (Fuchs, Fuchs, & Compton, 2012).

Demands for increased accountability in schools have prompted many educators to turn to systems that provide frameworks for data-based decision making that support service delivery models across general, remedial and special education settings (Fuchs & Vaughn, 2012; Stecker, Lembke, & Foegen, 2008). One example of such a system is Response to Intervention (RTI) in which teachers use progress-monitoring data to identify and target students who are not performing satisfactorily and then to track their growth during the provision of evidence based instructional interventions (Fuchs & Fuchs, 2007). In an age of accountability, data-based decision making and frequent progress monitoring has become the law of the land. Such educational approaches create the need for scientifically validated assessment measures that allow for screening of academic skill deficits, track growth during the provision of targeted interventions, and demonstrate that sufficient progress towards academic standards is met (Fuchs, Mock, Morgan, & Young, 2003; Stecker et al., 2008; Zirkel & Thomas, 2010) . Valid and reliable assessment measures are essential in today's era of increased accountability in education.

Instrument Validation

In order to validate a measure with standards that are considered acceptable to a scientific community, a mutually agreed upon set of criteria are needed. Within the field of educational and psychological assessment there are commonly accepted procedures for the examination of assessment measures which have been set forth by a joint committee of the American Education Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement Education (NCME) (2014). The agreed upon criteria can be found in The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), here on out referred to as *the Standards.*

Measurement is a broad process for quantifying behavior. Allen and Yen (2002) define measurement as a method of assigning numbers to individuals under study in a systematic way as a means of representing properties, characteristics, qualities, and behaviors of the individual. Furthermore, "numbers are assigned to the individuals according to a carefully prescribed, repeatable process" (Allen & Yen, 2002, p. 2). Testing is a specific and more formal type of measurement, which is considered to be a systematic procedure for collecting samples of behavior that have been identified by an operational definition within a particular construct domain (Crocker & Algina, 1986).

Classic test theory. Classical Test Theory (CTT) comprises a set of concepts and techniques that serve as the basis for many measurement instruments, and it stands as a reference point for more recent measurement methods, such as Item Response Theory (DeVellis, 2006). Researchers are often interested in phenomena that are not directly observable and therefore proxy indicators must be used. CTT provides direction and methods to evaluate how successful these proxy indicators are at estimating the levels of unobservable variables of interest (Allen &

Yen, 2002). CTT, as it is known today, has its roots in the work of Spearman and is based on the ideas of a true score model (Spearman, 1904; Traub, 1997). Later, CTT was further developed by Novick (1966) with his defense of the principles of CTT. Later texts continued to expand, explain and document the principles, theories, and applications of CTT (Allen & Yen, 2002; Crocker & Algina, 1986; Lord & Novick, 1968). CTT is founded on the idea that measurement error, a random latent variable, is a part or component of the observed score (Traub, 1997). Therefore an individual's True Score, or their theoretical true ability or performance on a measure, is made up of both the individual's observed score on a measure and measurement error (Allen & Yen, 2002).

$$X = T + E \text{ (observed score = true score + error)}$$

As noted, CTT provides the foundation for researchers to evaluate the reliability and validity of assessment measures.

Reliability. Reliability is the consistency with which a variable can be measured (Meyer, 2010). In the fields of psychology and education, an assessment instrument's reliability is most often indicated as a correlation coefficient (Weiner & Greene, 2008). Allen and Yen (2002) state that there are several ways of defining reliability, citing the example that if a test's observed scores are highly correlated with its true scores, then it would be considered reliable. Researchers often consider a reliability coefficient of at least .75 as the minimum necessary correlation to indicate that a variable is being assessed reliably (Weiner & Greene, 2008). As the reliability of a test increases, the error-score variance for the test becomes relatively smaller, which means that the smaller the error variance then the closer an individual's observed score is to his or her true score (Allen & Yen, 2002).

Internal consistency. A test's internal consistency is a measure for the homogeneity among the items it contains and is reliant upon the average intercorrelation among all the items on the test (Weiner & Greene, 2008). Internal consistency is a method of reliability analysis that requires only one administration of the assessment tool, and thus avoids problems that accompany repeated testing, such as test/retest reliability (Allen & Yen, 2002). Furthermore, the technique is useful because it solved the issues of estimating reliability with two-part split halves of a test and needing to decide which split was best for accurately estimating reliability (Meyer, 2010). Cronbach's alpha (Cronbach, 1951) is a measure of internal consistency and is a commonly used measure of reliability in the social sciences. Cronbach's alpha "describes the reliability of a sum (or average) of q measurements where the q measurements may represent q raters, occasions, alternative forms, or questionnaire/test items" (Bonett & Wright, 2015, p. 3). More simply put, Cronbach's alpha "tells you the degree to which the items are interrelated" (Johnson & Christensen, 2012, p.170).

Validity. Validity is the degree to which a set of scores represent the underlying construct that they are intending to measure (Allen & Yen, 2002). During the evaluation of an assessment measure, validity should be the most important consideration (AERA, APA, & NCME, 2014). According to Messick (1989) "validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (p.13). A broader definition of validity is that it "is an inductive summary of both the existing evidence for and the potential consequences of score interpretation and use" (Messick, 1989, p. 13). The *Standards* (AERA, APA, & NCME, 2014) emphasize the key aspect to evaluating validity is the degree to which evidence supports both the theory behind the use of a measure and

the intended use and interpretation of a test's meaning. Finally, the *Standards* state that the process of validation must involve accumulating evidence for a sound scientific basis of the proposed score interpretations (AERA, APA, & NCME, 2014).

According to Kane (2004) investigating the validity of an assessment measure can inform whether a tool is providing accurate, sound information about an individual or group being assessed and whether the basis for decisions being made with that information is sound. An approach to establishing validity is an argument-based method that has been supported by Kane (2001, 2004, 2006), who built upon previous work by Cronbach (1988), which articulated methods for verifying the meaning of scores from assessment tools. According to Kane (2006), in order to validate a given interpretation of a test, or the particular use of test scores, an evaluation of the interpretation or use is required. Therefore, the evidence needed to defend or assess the degree of validity depends upon the claim being made. Consequently, a validity claim needs a clear statement of the proposed use and interpretation of scores (Kane, 2006).

Conducting validation work requires two types of arguments, interpretive arguments and validity arguments (Kane, 2001; 2006). Interpretive arguments specify "the proposed interpretations and uses of test results by laying out the network of inferences and assumptions leading from the observed performances to the conclusions and decisions based on the performances" (Kane, 2006, p. 23). A validity argument provides an evaluation of an interpretive argument (Cronbach, 1988). Furthermore, Cronbach states an "argument must link concepts, evidence, social and personal consequences, and values" (p. 4). He elaborates further stating that when creating an argument for validity one must "understand what generates test scores and why they relate to antecedents and consequents, and to understand also the context of test use" (p. 4). Kane's approach to validity suggests that in order to claim that a proposed

interpretation or use of an assessment is valid, one must produce an interpretive argument that is coherent, with reasonable inferences, and that its assumptions are plausible. Furthermore, he suggests that the analysis should make clear to the greatest extent possible an argued construction of reality that explains the implicit value of the test and its application (Kane, 2006,).

In the fields of education and psychology valid and reliable measures are needed for the assessment of academic deficits, such as reading difficulties (APA et al., 2014). Children with developmental disabilities, such as ASD, are often not reflected in the norms used to validate the use of measures (Markus & Borsboom, 2013). In order to make accurate instructional decisions, assessment measures that are valid and reliable for use with minority populations are a necessary component of the education system (Fuchs, Compton, Fuchs, Bouton, & Caffrey, 2011).

Progress monitoring. Progress monitoring is a method of assessment which can aid in informing educational practices and producing better educational outcomes (Stecker et al., 2008). According to the National Center on Student Progress Monitoring "progress monitoring is defined as a scientifically based practice that is used to assess students' academic and/or behavior performance and evaluate the effectiveness of instruction or intervention" (2008). Progress monitoring requires frequent and repeated collection of data on student performance in academics, social behavior, or both. The data from progress monitoring allow teachers, school psychologists, and other educational professionals to make data-based decisions about a student's response to intervention.

To be useful in formative evaluation, progress-monitoring tools must meet technical adequacy standards for reliability and validity, must be sensitive to short-term changes in performance, and must be time-efficient so teachers can monitor student progress frequently.

When conducted correctly and systematically, progress monitoring promotes desirable outcomes for both staff and students (APA et al., 2014; Foegen, Jiban, & Deno, 2007; Fuchs, 2004). Outcomes of effective progress monitoring include (i) data-based decision-making, (ii) better emotional and behavioral functioning, (iii) greater accountability via documentation of student progress, and (iv) more efficient and effective communication between school staff, parents, and students (Stecker et al., 2008). Research has shown the importance of measuring student progress objectively and using obtained data to determine when and how to adjust academic or behavior support programs (Fuchs & Vaughn, 2012). This research indicates that frequent measurement and responsive use of that information in instructional decision making can enhance teacher planning and student outcomes (Coyne, Kame'enui, & Simmons, 2004; Fuchs et al., 2012; Sandall, Schwartz, & Lacroix, 2004). Finally, studies have demonstrated that progress monitoring is important when intervening on students' oral language and reading skills (Bowyer-Crane et al., 2008; Compton, Fuchs, Fuchs, & Bryant, 2006; Haager & Windmueller, 2001).

Reading

Reading is a highly valued skill that is important for social and economic advancement and necessary for general success in society (Griffin, Burns, & Snow, 1998). Whitehurst and Lonigan (2001) describe learning to read as a key milestone for children living in a literate society. The ability to read provides a critical foundation for a child's academic success throughout life, and fluent reading is critical to functioning in our modern society. Reading is a complex activity with many component skills and requisite knowledge (Adams, 1990) and, according to Hiebert (2009), reading material in modern society now exceeds a twelfth grade reading level. Unfortunately, more than one-third of all children experience significant difficulty when learning to read (Adams, 1990; Shaywitz, Escobar, Shaywitz, Fletcher, & Makugh, 1994).

It is critical that we understand how to teach all individuals in our society to read and write to sufficient standards.

Hammill (2004) states that a "good reader" is a student who can successfully engage in four subcomponents of reading. First, they can associate speech sounds with letters. Second they can pronounce printed words. Third, they gain meaning from print, suggesting they understand the content of what was read. Finally, they also read orally and silently with sufficient accuracy and speed to functionally extract meaning from a passage. These four subcomponents are often referred to as the "core" of what it takes to be a good reader.

Authors differ in how they parse and describe the subcomponents of the complex reading process. The report of the National Reading Panel (2000) identifies five important elements of reading instruction: phonemic awareness, phonics, fluency, vocabulary, and comprehension. Solid reading instruction must also target print awareness and letter knowledge (Foorman & Torgesen, 2001; Foulin, 2005; Storch & Whitehurst, 2002). An alternative theory about reading suggests that becoming a skilled reader requires the knowledge and ability to weave together language comprehension and word recognition (Neuman & Dickinson, 2001). In this approach, language comprehension includes background knowledge, vocabulary, language structures, verbal reasoning, and literacy knowledge; whereas word recognition includes phonological awareness, decoding, and sight word recognition (Brassard & Boehm, 2007; Neuman & Dickinson, 2001). Still others suggest that the knowledge, skills, strategies, and abilities that are needed to be a proficient reader with successful comprehension will change gradually over the course of development (Oakhill & Cain, 2007). Regardless of the theoretical approach to how children learn to read, Whitehurst and Lonigan (1998) assert that even from the earliest stages

the process of reading is driven by the extraction of meaning, which provides the drive to develop the ability to read and write, otherwise known as literacy.

Simple View of Reading. The Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990) describes the ability to read as a combination of both decoding (i.e. word recognition) and oral language skills. The Simple View (Gough & Tunmer, 1986) is a helpful model to explain and elaborate upon many common challenges individuals face when learning to read (Gough & Tunmer). Ricketts (2011) combined the Simple View and research by Catts, et al. (2003) to create a multi-dimensional approach to skilled reading. An illustration and expansion on the Simple View of Reading can be found in Figure 1. The Simple View can be represented as a field broken into four quadrants. In quadrant A, an individual has poor word recognition, but intact language comprehension. In quadrant B, an individual has both intact word recognition and language comprehension. In quadrant C, an individual would have both poor word recognition and poor language comprehension. Finally, in quadrant D, an individual would have intact word recognition, but poor language comprehension. Furthermore, Gough's model pairs well with other theoretical frameworks.

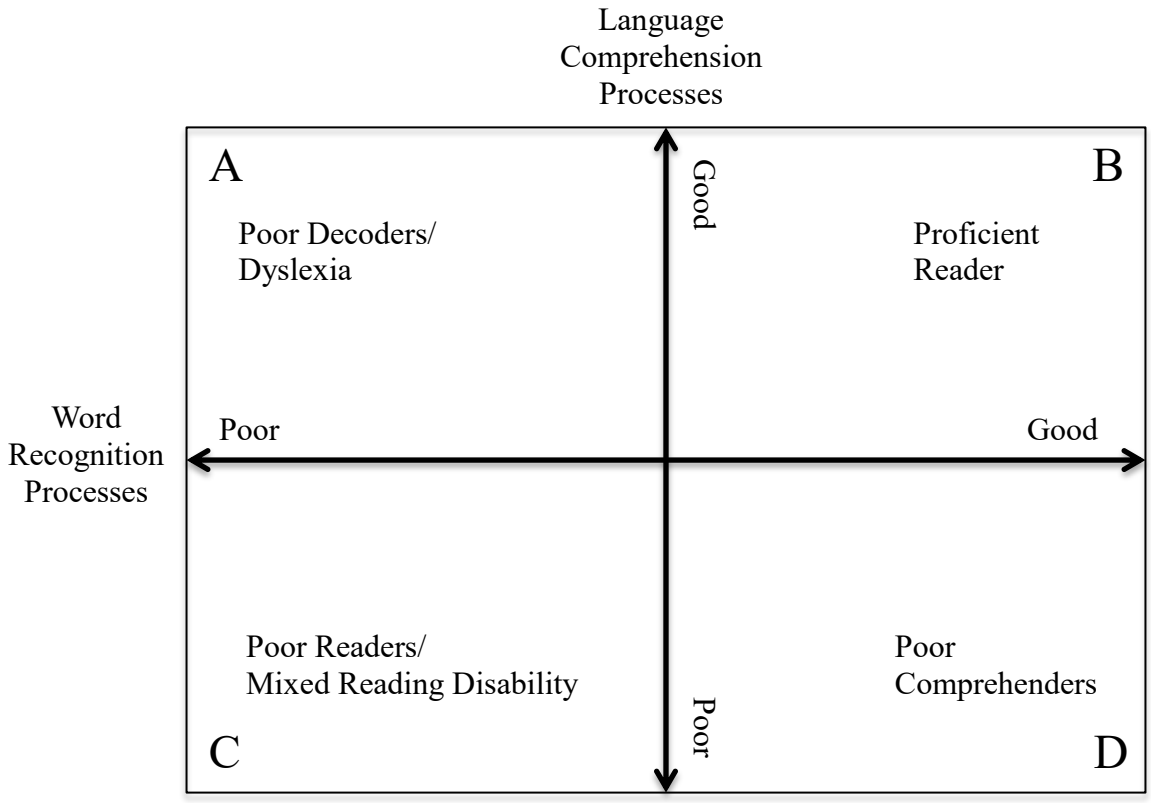

Figure 1. Simple View of Reading. This figure was adapted from Ricketts' (2011) work and illustrates the Simple View of Reading (Gough, 1986).

The Simple View dovetails effectively with Whitehurst and Lonigan's (1998) breakdown of emergent literacy as being a set of outside-in and inside-out skills. Gough (1986) suggested that the final goal of reading is reading compression. The two main driving factors that lead to good reading comprehension are decoding and listening comprehension. Pennington and Bishop (2009) expand upon Gough's view of decoding, breaking the construct into component skills of phonological awareness, orthographic coding, and rapid serial naming. Combined, these skills lead to fluent printed word recognition (i.e. decoding). Listening comprehension incorporates

the understanding of linguistic information and combines vocabulary, morphosyntax, and pragmatics.

Common reading disorders that map onto the Simple View are dyslexia, mixed reading disability (i.e. poor readers who struggle with both decoding and comprehension) and poor comprehenders (Davidson & Weismer, 2014). Individuals with dyslexia, or poor decoding skills, map onto quadrant A. Hulme and Snowling (2009) estimate that 3 to 6% of the general population deal with dyslexia. Proficient readers who are able both to easily recognize words and comprehend language would map onto quadrant B. Individuals who struggle with both word recognition and language comprehension would map onto quadrant C. Individuals in quadrant C are often categorized as poor readers, or as having a mixed reading disability (Davidson & Weismer, 2014; Gough & Tunmer, 1986) Finally, individuals who struggle with comprehending text, without any apparent difficulty decoding written text, would map onto quadrant D. It is estimated that 7 to 10% of the general population would fall into this category (Clarke, Snowling, Truelove, & Hulme, 2010; Nation, Cocksey, Taylor, & Bishop, 2010).

Emergent literacy. Early literacy, or emergent literacy, builds the foundation for a child's journey into reading (Lonigan, Burgess, & Anthony, 2000). According to Whitehurst and Lonigan (1998) early literacy encompasses the skills, knowledge, and attitudes that are developed *before* an individual reads and writes. They further state that "the acquisition of literacy is best conceptualized as a developmental continuum, with its origins early in the life of a child, rather than an all-or-nothing phenomenon that begins when children start school" (Whitehurst & Lonigan, 1998, p. 848). Evidence supports a substantial connection between early reading skills and an individual's later ability to read (Duncan et al., 2007; Wagner et al., 1997).

The concept of emergent literacy differs from more traditional ideas of reading, otherwise known as conventional literacy (Whitehurst & Lonigan, 2001). The National Early Literacy Panel (2008) (NELP) identifies conventional literacy as a set of skills that includes decoding, oral reading fluency, reading comprehension, writing and spelling. The NELP also identifies six early literacy skills, which are precursor skills that have predictive relationships with conventional literacy skills. These six early literacy variables are alphabet knowledge, phonological awareness, Rapid Automatic Naming (RAN) of letters or digits, RAN of objects or colors, writing and writing one's own name, and phonologic memory (National Early Literacy Panel, 2008). Furthermore, NELP identified five additional early literacy skills, which are somewhat correlated with later literacy; these are concepts about print, print knowledge, reading readiness, oral language, and visual processing. For developing readers, these eleven variables were found to be consistently predictive of literacy achievement in preschool and kindergarten However, this set of 11 early skills was found to be more predictive for learners in late kindergarten and early first grade than for learners' later literacy development in grades four through six (NELP).

*Domains of emergent literacy***.** Components of emergent literacy can be viewed as two separate domains. Whitehurst and Lonigan (1998) proposed that both emergent and conventional literacy are two interdependent sets of processes and skills. Broadly, they described the two areas as outside-in and inside-out units (Whitehurst & Lonigan, 1998). Outside-in units include the narrative, conceptual, and semantic context of read information. Inside-out units include skills to decode letters in a passage into phonological units that have meaning to the individual. Such skills require a knowledge of print units such as graphemes, sound units such as phonemes, and the cognitive processes needed to organize these units into

words and sentences (Whitehurst & Lonigan, 1998). A longitudinal study by Storch and Whitehurst (2002) found that reading ability in early elementary school is primarily determined by children's level of print knowledge and phonological awareness, variables that reflect the inside-out skills an individual has when entering kindergarten. The NELP (2008) also identified certain aspects that relate to inside-out units (i.e. phonological processing and phonological awareness) as being causally connected to later literacy achievement. As early literacy progresses, the instructional methods that involve outside-in and inside-out skills become separated in kindergarten and first grade at which time the focus of learning to read is primarily on decoding words or inside-out units. As learning experiences move in the second grade to outside-in domains of literacy, the instructional focus for typically developing learners shifts toward comprehension (Whitehurst & Lonigan, 1998). In sum, the experiences, opportunities, and instruction that support the development of inside-out and outside-in skills differ from each other, and an individual's early differences between these skill sets are stable across time as their ability to read continues to develop (Byrne, Freebody, & Gates, 1992; Wagner, Torgesen, & Rashotte, 1994; Whitehurst & Lonigan, 1998).

Phonological processing. Phonological processing is an auditory processing skill that is made up of three distinct yet related areas: phonological awareness, phonologic recoding in lexical access, and phonetic recoding in working memory (Wagner & Torgesen, 1987). According to Adams (1990) phonological processing is important not just for spoken language but also for reading print. She posits that phonological processing optimizes reading performance by providing a redundant word processing route that is critical to maintain both speed and accuracy of word recognition during productive reading. It also expands the amount

of cognitive memory a reader can use when placing words in the context of a passage for text comprehension (Adams, 1990).

Phonological awareness. Phonological Awareness (PA) is defined as an awareness of the sound structure of language (Wagner & Torgesen, 1987). Wagner & Torgesen describe PA as also including the conscious ability to detect and manipulate sound in spoken language (i.e., to move, combine, and delete sound parts) at multiple levels of abstraction and to access the sound structures of spoken language. The process of building PA prepares emerging readers for later instruction in reading, including components such as phonics, word analysis, and spelling (Chard & Dickson, 1999; Chard, 1998). Over time scholars have argued about how best to represent and break down PA, positing that it could be represented as a one, two or four factor model (National Early Literacy Panel, 2008; Stanovich, Cunningham, & Cranner, 1984; Yopp, 1988).

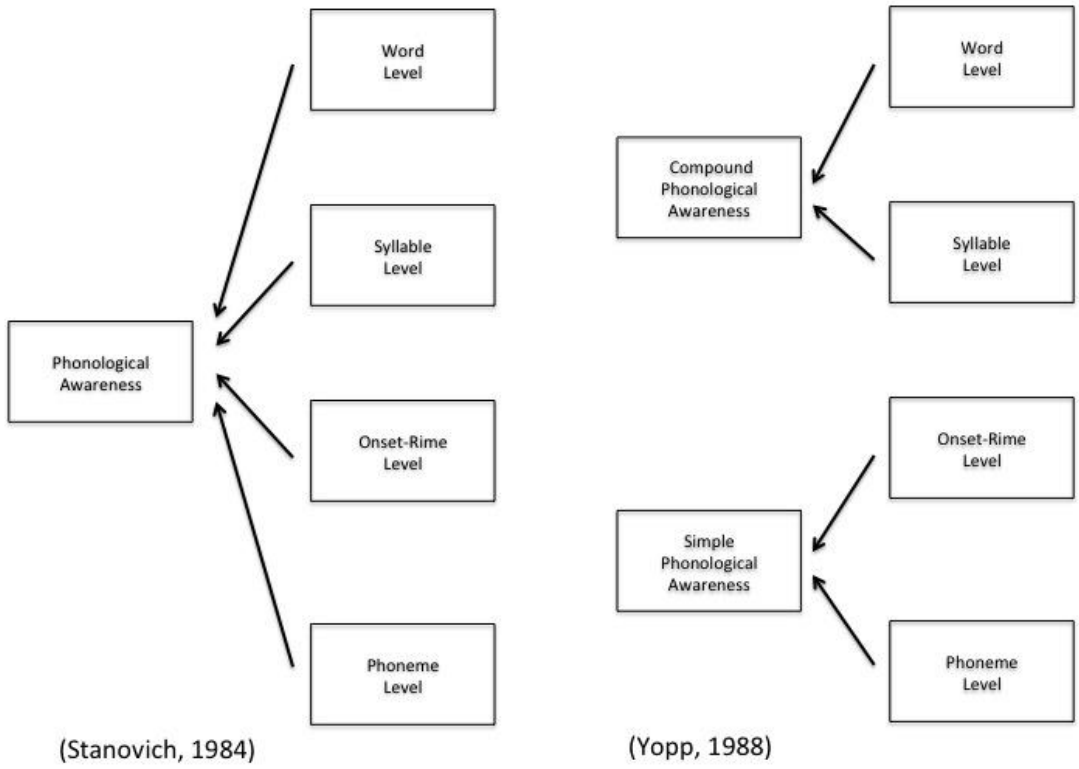

Figure 2. Phonological Awareness Represented as One and Two Factor Models.

PA includes the ability to detect, isolate, manipulate, blend, and segment language at four different levels which are: (a) word level, (b) syllable level, (c) intra-syllabic level (onsets and rimes), and (d) phoneme level (National Early Literacy Panel, 2008). Young children develop PA in a progressive manner moving from larger units (words) toward shorter and more abstract parts of speech (Wagner et al., 1997). Children become sensitive to word level first, followed by syllables, onset rimes, and then phonemes (Whitehurst & Lonigan, 1998). Chard and Dickson (1999) consider an individual having a general understanding at all four of these levels to be phonologically aware.

*Phonological awareness and reading***.** An over-simplified view of PA's vital role in learning to read is that a beginning reader must learn that symbols correspond to units of speech, and that units of speech are represented as phonemes (Crowder, 1982). When learning to read new words, an individual must segment letter strings into units, which correspond to individual sounds (i.e. phonemes). These individual sounds are then blended together in order to produce and pronounce a word (Wagner & Torgesen, 1987). Gillon (2004) considers a student's PA to be the best single predictor of reading performance in most children. A meta-analysis by the NELP (2008) found that PA had medium to large predictive correlations with decoding and reading comprehension at .40 and .44 respectively. Four phonological tasks that are considered to be significant predictors of reading ability are: discriminating rhymes, segmenting syllables, isolating phonemes, and blending phonemes (Bryant, MacLean, Bradley, & Crossland, 1990; Goswami & Bryant, 1990).

An examination of the relationship between phonological awareness and reading provided evidence for seven important conclusions (National Reading Panel, 2000). The first of these findings is that phonological awareness is directly related to reading ability. Second,

phonological awareness has a reciprocal relationship with decoding, however PA is a skill that precedes the acquisition of skills to decode written text. Third, phonological awareness has been shown to be a reliable predictor of later reading ability. Fourth, deficits in phonological awareness are usually associated with deficits in reading. Fifth, early language experiences play an important role in the development of phonological awareness. Sixth, guided early intervention can promote the development of phonological awareness. Finally, an improvement in PA often results in improvements in reading ability (National Reading Panel, 2000). Because phonological awareness supports later reading success, Phillips and Piasta (2013) call PA a foundational early literacy skill.

Most research investigating the effects of PA and phonological skills focuses on the relationship between decoding and word reading (Oakhill & Cain, 2007). In spite of the NELP's (2008) finding that phonological awareness is moderately to strongly correlated with later comprehension, Oakhill and Cain's (2007) review of the literature suggested that studies which have linked the two have produced mixed results. A two-year longitudinal study of 90 British children aged 4 to 5-years-old found that comprehension was best predicted by prior word recognition skills, vocabulary knowledge, and grammatical skills (Muter, Hulme, Snowling, & Stevenson, 2004), all of which are abilities that Whitehurst and Lonigan (1998) would consider outside-in skills. Other studies have found that the relationship between comprehension and contributing skills changes with age (Rupley & Wilson, 1997). Rupley and Willson (1997) found that in early schooling, grades 2 to 4, phonologic knowledge and prior knowledge of the topic were the primary factors contributing to reading comprehension. By grades 5 and 6, a student's successful comprehension of reading content was better predicted by knowledge of reading strategies, such as metacognitive and predictive strategies (Rupley & Willson). These

findings suggest that reading is a developmental continuum, rather than a static construct, with different skills and abilities presenting different levels of contribution as an individual grows up and becomes a more proficient reader. However, some young children, such as those diagnosed with ASD, don't follow this trajectory of acquiring reading skills. There remains a gap in the literature in trying to identify and understand how children with ASD acquire reading skills.

Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a pervasive neurological disorder with a wide range of impacts, outcomes, and varying levels of severity (American Psychiatric Association, 2013; Bernier & Gerdts, 2010; Billstedt, Gillberg, & Gillberg, 2005). ASD typically produces unusual patterns of development that begin in infancy and early childhood years. This atypical pattern of development is characterized by deficits in two areas: (i) social communication and social interaction; and (ii) restricted, repetitive patterns of behavior, interests and activities, and sensory differences (American Psychiatric Association, 2013). The prevalence rates are increasing with debate among experts as to whether there is a true increase in numbers of cases or if the profession is simply getting better at recognizing and diagnosing the disorder (Harvard Health Publications, 2010). Furthermore, it is unclear if a changing definition of ASD has created a moving target, or if the real answer lies within some combination of factors. Although debate on the exact rates of ASD continues, as children with ASD grow into adults more and more people will be impacted directly or indirectly by a friend, loved one, or coworker with ASD (Montes & Halterman, 2008).

ASD is a neurological disorder which, with no known cure, presents as a lifelong disability (National Institute of Child Health and Human Development, 2005). ASD has been increasingly framed in a new and different perspective as neurological difference. (Kapp,

Gillespie-Lynch, Sherman, & Hutman, 2012). Referring to ASD as a neurological difference, rather than a neurological disorder, has the power to normalize the condition and frames it as a positive identity that requires no cure (Kapp et al.). This change in perspective is a core difference from the medical model's treatment of the disorder. A deficit vs. difference model may appear to be merely semantics, but is argued to be more mindful of human dignity and to promote more subjective levels of human wellbeing (Kapp et al.). Small everyday examples include a use of person first language such as people with autism, rather than autistic people. Regardless of perception, if the individual is provided intervention, skill building, and environmental adaptations and accommodations, impacts of the disorder can be ameliorated. Furthermore, evidence indicates that early intervention is imperative for optimal outcomes (Stoelb et al., 2004).

The characteristics and behaviors that are typically exhibited by individuals with ASD are unique and fascinating (Bernier & Gerdts, 2010). Each individual with autism has a unique mix of symptoms with much variability from one individual with ASD to another. It is often considered a heterogeneous disorder with a variety of etiologies and multiple causes and courses of progression (Geschwind & Levitt, 2007). According to Webber and Scheuermann (2008), the most severe forms of the disorder present with symptoms in *all* of the following categories: communication and language deficits, cognitive, social, and sensory processing deficits, and stereotyped behavior. However, depending on severity of symptomatology, individuals with ASD may or may not present with challenges in each area. As understanding of the disorder has evolved, researchers and professionals who determine diagnostic criteria (American Psychiatric Association, 2013) have focused on symptoms in the two core areas of reciprocal social interaction and communication and of restrictive and repetitive behaviors.

Reciprocal social interaction and communication. Reciprocal social interaction and communication is the first of two areas of difficulty which characterize ASD (American Psychiatric Association, 2013). This encompasses greetings, emotional sharing and reciprocity, offering and seeking comfort to or from another person, and the development of intense meaningful friendships (Couteur et al., 1989). Other researchers describe the presentation in this domain as a general lack of motivation to communicate, with some individuals with ASD having little to no interest in interacting with people at all (Webber & Scheuermann, 2008). Common everyday social impairments listed by Sansosti, Powell-Smith, & Cowan (2010) include the following: difficulty interacting with others even if the individual demonstrates a desire for friends, a broad difficulty maintaining social relationships, challenges following social rules/norms, and a struggle to understand facial expressions, gestures, or vocal tone. Additionally, there are often deficits in understanding another person's perspective, also known as Theory of Mind (Baron-Cohen, 1995). Finally, individuals may occasionally, or even frequently, exhibit culturally inappropriate social exchanges and behaviors (Sansosti et al., 2010).

In addition to difficulties with reciprocal social interaction there may be differences in communication and language. This includes features such as social usage, style of conversational interchange, the use of idiosyncratic language such as language with private meanings that make sense only to those familiar with the situation in which the phrase originated, and stereotyped utterances including abnormal or excessive repetition of a phrase (Couteur et al., 1989). Common communication difficulties listed by Sansoti et al. (2010) are: challenges in initiating or maintaining conversation; a tendency to dominate conversations and talk excessively about his/her own topic of interest; challenges with turn taking; and, failure to

respond with consistency to comments, questions, or other social bids from conversational partners. Furthermore, individuals with ASD often use inappropriate register (the style of language used for a particular purpose or in a particular setting), such as using very formal language when talking within an informal social context or using vulgar language in a formal situation. People with ASD may also demonstrate odd prosody, or patterns of stress, rhythm, and intonation in their speech. Some individuals with the disorder interpret jokes, sarcasm, or idioms very literally. Finally, they may struggle to comprehend implied instructions or meanings.

Repetitive and restricted behaviors. The second core area of symptomology is repetitive, restricted, and stereotyped behaviors (American Psychiatric Association, 2013). Examples include unusual preoccupations, strong attachments, rituals, and unusual sensory interests (Couteur et al., 1989). Common unusual behaviors that Sansoti et al. (2010) suggest include: an engagement in rigid motor mannerisms, such as hand flapping or pacing, and an unusual insistence on adherence to routines or rituals. Individuals may display a greater interest in parts of objects than in the objects themselves, as the child who only spins wheels on vehicles but doesn't push/play with the vehicles as toys. They may have developed substantial factual knowledge of a relatively narrow topic of interest, often referred to as a restricted interest (Eigsti, de Marchena, Schuh, & Kelley, 2011). Finally, individuals may focus the majority of their social interactions on one of their identified topics of interest. Such children have been referred to as 'little professors' as many of their conversations may delve into narrow but well researched and sometimes obscure topics (Donnellan, 1985; Tager-Flusberg & Caronna, 2007).

Prevalence. Over the course of recent history there has been a drastic increase in diagnosed rates of ASD to the current rate of 1 in 68 people (National Center for Health

Statistics, 2013; CDC, 2016). Although studies have cited different ASD prevalence rates, most studies seem to agree that the rates of Autism are increasing, though it is challenging to untangle what is leading to increased levels of identification (Rice et al., 2012). Studies have shown a statistical increase in prevalence since the 1960s, however some researchers argue this is due to increased public awareness and diagnostic criteria which have become broader and more inclusive (Harvard Health Publications, 2010).

When ASD was first studied, it was considered to be a low-incidence disability and had a prevalence rate of 4-6 per 10,000, or about 1 in 1,600 individuals (Lotter, 1966). Although there is disagreement about exact current prevalence rates, a review of published studies indicates that ASD rates have increased drastically (Christensen et al., 2016; Williams, Mellis, & Peat, 2005). For instance Fombonne and colleagues (2003) suggested that rates for ASD were 3 to 4 times higher than they were 30 years earlier. In recent years the estimates of the occurrence of ASD in the United States have exploded (Newschaffer, Falb, & Gurney, 2005). In 2007, the CDC estimated the prevalence rate to be about 1 in every 150 children (2007). Just a few years later, a new study cited a rate of 1 in every 110 children (CDC, 2009). Finally, the most recent studies suggest that the rate is now closer to 1 in 68 children being identified with ASD (CDC, 2014). Researchers are uncertain as to whether this is due to a broader definition of diagnostic criteria, or simply due to better efforts in diagnosing (Harvard Health Publications, 2010; Presmanes-Hill, Zuckerman, & Fombonne, 2015). While there is debate, an increase of the prevalence is not in doubt and some believe the true reason lies within a combination of a broader definition and better diagnostics (Matson & Kozlowski, 2011). The American Academy of Pediatrics suggest that 1 to 2 percent of children are currently being diagnosed with ASD (Harrington & Allen, 2014) and this is roughly in line with the most recent data to date provided by the CDC

(Christensen et al., 2016), which estimates that current rates of diagnosis of ASD among school-aged children are 1 in 68 (14.6 per 1,000).

Academic difficulties. Individuals with ASD have widely varied academic performance (Minshew, Goldstein, Taylor, & Siegel, 1994; Nation et al., 2006). Learning and academic outcomes range from severely impaired to levels of exceptional performance (Griswold, Barnhill, Myles, Hagiwara, & Simpson, 2002). A study by Estes and colleagues (2011), sampled 30 high functioning 9-year old children with ASD to examine discrepancies between academic achievement and intellectual ability. They found that 90% of their sample demonstrated a discrepancy between predicted achievement based on cognitive assessment of their intellectual functioning and actual performance on academic measures in at least one of several academic domains, such as word reading, spelling, or basic math skills (Estes, Rivera, Bryan, Cali, & Dawson, 2011). In other words, unlike some other disorders or disabilities, i.e. a specific learning disability, a diagnosis of ASD does not necessarily predict where a child's academic performance will fall.

Reading and children with autism spectrum disorders. Elementary-aged children and youth with Autism Spectrum Disorders (ASD) are a very heterogeneous population which complicates examining their ability to read (Nation, Clarke, Wright, & Williams, 2006). Children with an ASD may or may not have several other comorbid diagnoses, such as an intellectual disability or accompanying language impairment. Therefore when grouped together for the purpose of examining reading abilities, children with ASD do not fit into homogeneous categories as easily as do children with other diagnoses such as dyslexia or speech and language impairment (Griswold et al., 2002; Minshew et al., 1994; Nation et al., 2006).

A study by Nation et al. (2006) is credited with having one of the most heterogeneous populations of individuals with ASD, meaning participants were from a range of levels regarding cognitive, linguistic and reading abilities. Many other studies that investigated the academic abilities of children with ASD have focused on high functioning individuals, excluding their lower functioning peers (Griswold et al., 2002; Minshew et al., 1994). In Nation's (2006) study of 41 participants ages 6 to 15, 78% of participants had measurable reading abilities, meaning they could read aloud at least one word out of context. Furthermore, 65% of her sample demonstrated impaired reading comprehension and a third of her total sample demonstrated very severe impairment in reading comprehension. In addition, a notable portion of participants also struggled with decoding. Of the 32 participants that could read 42% showed moderate impairment in measures of non-word reading and 22% demonstrated severe impairment. These findings indicate that children with ASD cannot be neatly categorized into one group of reading disorders. Some individuals with reading disorders are poor comprehends, others are poor decoders, and some individuals struggle with both (Nation et al., 2006).

In a follow up study that built upon Nation and colleagues' (2006) work, Davidson and Weismer (2014) investigated reading profiles of very young, higher functioning children with ASD to examine whether an early reading profile of strong print-related skills and lower language/vocabulary skills could be replicated during early phases of reading development. Of the 101 participants in the study, 62% of the sample were found to fit this profile, which corresponds to quadrant D or poor comprehenders, in Figure 1. Furthermore, using a latent profile analysis Davidson (2014) found that 31% grouped in a profile that corresponded to quadrant C, which indicates challenges with both decoding and comprehension.

Other researchers have also documented the emergent literacy skills of children with ASD. A relatively small study conducted by Gabig (2010) included 14 students with ASD and 10 age matched typically developing peers with ages ranging from five to seven years. Surprisingly, her study found no relationship between PA tasks and measures of word reading in children with ASD. In her sample, children with ASD demonstrated adequate ability in single word reading despite having below average PA skills. Another study investigated the emergent literacy skills of 41 children with ASD ranging in age from four to eight years old (Lanter et al., 2012). These students were grouped by language development level, as typical language individuals, students with mild to moderate language impairments, and those with severe language impairments. The study found that emergent literacy skills correlated with language development. These results from Lanter's (2012) study were consistent with the link between language skills and emergent literacy that had earlier emerged in a study of typically developing children (National Institute of Child Health and Human Development, 2005). Another study conducted by Dynia et al. (2014) found that after controlling for level of language skills, ASD status was a significant predictor for skill levels regarding alphabet knowledge and print-concept knowledge. They found that children with ASD had significantly higher alphabet knowledge than their Typically Developing (TD) peers and significantly lower print-concept knowledge compared to TD peers. However, ASD status was not a significant predictor for either definitional vocabulary or phonological awareness. Furthermore, they found that children with ASD had significantly lower levels of print interest than their TD peers after controlling for language skills. The relative weaknesses shown by children with ASD in respect to print-concept knowledge is an important finding in regards to later literacy outcomes. According to

Hammill (2004) and the NELP (2008), print-concept knowledge is a consistent predictor of later reading achievement.

Scholarly work indicates various types of reading difficulties exist amongst children with ASD (Davidson & Weismer, 2014; Gabig, 2010; Nation et al., 2006). In order to identify, treat, and track reading and language intervention, valid and reliable assessment instruments are needed (APA, NCME, Joint Committee on Standards for Educational and Psychological Testing, & AERA, 2014; Chard & Dickson, 1999; Stecker, Lembke, & Foegen, 2008)

Purpose of the Study

Statement of the problem. In an age of accountability, it is imperative that educators have valid and reliable measures of student achievement and progress. It is well documented that some individuals with ASD have difficulties with reading (American Psychiatric Association, 2013; Estes et al., 2011; Griswold et al., 2002; Koritsas & Iacono, 2011; Minshew et al., 1994; Nation, Clarke, Wright, & Williams, 2006; Sansosti et al., 2010; VanTilborg, Segers, van Balkom, & Verhoeven, 2014). Students with ASD are often thought to mainly have difficulties with comprehension (Brown, Oram-Cardy, & Johnson, 2013; Gabig, 2010; Huemer & Mann, 2010; Nation & Norbury, 2005; Nation et al., 2006; Sansosti et al., 2010). However, within the context of the Simple View of Reading (refer to Figure 1), Davidson and Weismer's (2014) latent profile analysis found that 31% of their sample of children with ASD corresponded to quadrant C, which indicated challenges with both decoding and comprehension. Even though less than half of children with ASD appear to have this issue, Davidson and Weismer's findings indicate that some early readers with ASD may require PA intervention to ensure they become proficient readers. Thus, it is important that the assessment measures that are being used to identify academic performance problems among the ASD population be valid, reliable, and

effective measures of PA, which can then be subsequently used for educators to document and track progress of students. Such documentation can be used for accountability and to ensure the provision of high quality and effective intervention, so that children with ASD learn the skills they need to become proficient readers.

Theoretical orientation. Classical Test Theory (Allen & Yen, 2002; DeVellis, 2006) and Kane's argumentative validity framework (Kane, 2004, 2006, 2013) guide this study. CTT provides the foundational theories and methods for the evaluation of test measures and Kane's framework provides the structure to support the study's validity claims. An overview of Kane's framework for argument-based validation can be found in Figure 3.

Trait level	Description
Scoring	Scoring from observed performance to the observed score
	A1.1 The scoring rule is appropriate.
	A1.2 The scoring rule is applied as specified.
	A1.3 The scoring is free of bias.
	A1.4 The data fit any scaling model employed in scoring.
Generalization	Generalization from observed score to universe score
	A2.1 The sample of observation is representative of the universe of generalization.
	A2.2 The sample of observation is large enough to control random error.
Extrapolation	Extrapolation from universe score to target score
	A3.1 The universe score is related to the target score.
	A3.2 There are no systemic errors that are likely to undermine the extrapolation.
Implication	Implication from target score to verbal description
	A4.1 The implications associated with the trait are appropriate.
	A4.2 The properties of the observed scores support the implications associated with the trait label.

Figure 3. Interpretive Argument for a Trait Interpretation (Kane, 2006).

Construct definition. According to Kane (2006) a trait is the "disposition to behave or perform in some way in response to some kind of stimuli or tasks, under some range of circumstances" (Kane, 2006, p. 30). The trait under examination in my validity study is Phonological Awareness (PA), which is a metalinguistic task. Metalinguistic awareness entails

thinking about one's language, much like metacognition entails thinking about one's thinking. Wagner and Torgesen (1987) describe PA as the degree of awareness of the sound structure of language. Furthermore, PA is a construct which includes the skills of detecting rhymes, manipulating syllables, and manipulating individual phonemes by counting or deleting them (Whitehurst & Lonigan, 1998). The *Sound Beginning* Phonological Awareness Assessment (SBPAA) is broken into subscales that evaluate skills at the word level, syllable level, onset-rime level, and phoneme level (Lane & Pullen, 2004).

Research questions and hypotheses. The purpose of this study is to conduct an investigation of the psychometric properties of a phonological assessment measure used in the Preschool Autism Literacy Study (PALS; Hudson et al., 2017), a multi-year research project. PALS investigated the use of two emergent literacy interventions with preschoolers with ASD. The interventions are well established as being effective with typically developing readers. The project sought to determine what effect these same interventions would have for preschoolers with ASD. During PA intervention, the progress of participants' PA skills were monitored and assessed across multiple time points using the *Sound Beginning* Phonological Awareness Assessment (SBPAA) published by Lane and Pullen (2004). This measure did not include validity and reliability data for a subpopulation of typically developing preschoolers or for individuals with ASD. The need for an assessment of the validity and reliability of the measure with preschoolers is supported by standard 11.2 in the *Standards* (APA et al., 2014), which states "when a test is to be used for a purpose for which little or no documentation is available, the user is responsible for obtaining evidence for the test's validity and reliability for this purpose" (p.113).

Research question #1: What level of equivalency do form A and the alternate form B of the Sound Beginning Phonological Awareness Assessment (SBPAA) possess?

Hypothesis #1: It is hypothesized that forms A and B of the SBPAA will produce comparable results that will support equivalency.

Rationale #1: Equivalency between alternate assessment forms is essential so test scores are comparable. When conducting progress monitoring, it is important to have multiple forms of a measures that produce consistent scores (Stecker et al., 2008).

Research question #2: What is the internal consistency for the SBPAA when used with preschoolers with ASD?

Hypothesis #2: It is hypothesized that the SBPAA will produce scores that are within established limits of reliability for assessment measures.

Rationale #2: The internal consistency of the SBPAA is important because it is critical to know how well items on the assessment measure the same idea. Internal consistency is one of the most commonly used measures of reliability because it can be established from one administration (Henson, 2001).

Research question #3: What are the levels of item difficulty, discrimination, and discriminability for forms A and B of the SBPAA?

Hypothesis #3: It is hypothesized that an item analyses will provide support for each items' inclusion on this measure.

Rationale #3: Item level data is important to evaluate the value of questions in an assessment measure. Time is a valuable commodity in education, and if an item does not contribute meaningful or reliable information to an assessment the item should be revised or dropped.

Research question #4: What is the degree of convergent validity between the SBPAA and the TOPEL on a sample of preschool children with ASD?

Hypothesis #4: It is hypothesized that when compared to scores obtained using the Test of Preschool Early Literacy-TOPEL (Lonigan, Wagner, & Torgesen, 2007) the SBPAA will have acceptable levels of convergent validity.

Rationale #4: Measures of convergent validity are important because a comparison of the correlation coefficient between established tests strengthens assessment data gathered by the SBPAA. Convergent validity will be investigated because administrations of the SBPAA and TOPEL were not close enough in time to establish concurrent validity and not far enough apart to establish predictive validity.

Research question #5: To what degree is the SBPAA able to measure growth of phonological awareness skills in preschoolers with ASD?

Hypothesis #5: It is hypothesized that the SBPAA will be able to detect meaningful change when used as a progress-monitoring tool with the population drawn from the Preschool Autism Literacy Study, because the measure is aligned closely with lessons taught in the PA intervention subjects received.

Rationale # 5: The ability to detect change in skills is an essential component to a progress monitoring assessment, so a measure must be able to accurately detect change in order to inform intervention (Stecker et al., 2008).

Chapter III: Methods

Setting and Participants

Participant data for this secondary data analysis were a subset from the Preschool Autism Literacy Study (PALS), a multi-year study of students with Autism Spectrum Disorder (Hudson, et al., 2017). The participants in the PALS study were drawn from 32 schools across eight school districts in the Pacific Northwest. PALS investigated the immediate and long-term treatment effects of two early literacy interventions with preschool children with ASD. This study's participants were from the subset of PALS participants who received the Phonological Awareness treatment.

In 2013, 2014, and 2015, 151 students with ASD were recruited and their parents consented to their participation in the PALS project. Participants were randomly assigned to conditions across teachers. After attrition, the final PALS dataset included 47 students who received the Dialogic Reading (DR) treatment, 42 who received the Phonological Awareness (PA) treatment, and 44 who were assigned to the Business As Usual control condition. As previously stated, for the purpose of this analysis, data from 40 participants in the PA treatment condition were used in the analysis. Due to attrition two PA participants were dropped from the final data set.

Inclusion criteria. Inclusion criteria for PALS stated that a participant must be in their last year of preschool. Furthermore, they must have had an Individualized Education Plan and received services under the Individuals with Disabilities Education Act (IDEA) category of Autism. Finally, the participants needed to score at least a standard score of 55 on a measure of expressive language, the *Expressive One-Word Picture Vocabulary Test-IV* (EOWPVT-IV;

Martin & Brownell, 2011). All participants attended an integrated preschool and most of the students also received extended day services for ASD.

Demographic characteristics of the sample in this analysis are represented in Table 1. The majority of participants in the sample were male (n = 32, 80% of sample) and ages at pre-test ranged from 39 to 62 months (M = 55.25, SD = 4.79). The racial composition of the sample was diverse, with about half of the PA sample group identifying as part of a racial minority: 20 of the students were identified by their parents as being Caucasian, 8 identified as being from Asian heritage, 5 identified as Black, 3 identified as being Hispanic, and 4 identified as multiracial or other.

Table 1

Sample Descriptive Statistics

Characteristic	PA Intervention Group	
	n = 40	
	N	*%*
Male	32	80%
Minority	20	50%
Asian	8	20%
Black	5	13%
Hispanic	3	8%
Mixed/Other	4	10%
Mean Age at Pretest	55 mo.	

Note. Descriptive statistics of Phonological Awareness Intervention group.

Measures/Instruments

Phonological Awareness. Phonological Awareness (PA) was measured using two tools. In addition to the *Sound Beginning* Phonological Awareness Assessment (SBPAA), the PALS project made use of the *Test of Preschool Early Literacy*-TOPEL (Lonigan et al., 2007), which is a standardized assessment designed to identify preschoolers who are at risk for literacy problems that affect reading and writing. The TOPEL was administered to participants prior to

intervention and again after intervention. According to the test manual (Lonigan et al., 2007) internal consistency reliability measures for the three subtests ranged from .86 to .96 for their sample of 3 to 5-year-olds. Test-retest reliability for a one to two-week period ranged from .81 to .89. Concurrent validity for the subtests ranged from .59 to .77. The composite Emergent Literacy index produced an internal consistency reliability of .96.

The SBPAA is an early literacy resource and set of assessments assembled by Pullen, Lane, and Hayes (1999) and later published by Lane and Pullen (2004). For the purpose of the PALS study, the phonological awareness section of the assessment of metalinguistic abilities was used as a progress-monitoring tool during intervention. Over the course of the intervention two forms were used. Form A comprised questions from the original assessment (Lane & Pullen, 2004). Form B was created by the Principle Investigator of the PALS study and a doctoral student in special education, and it was designed to have similar items that matched the purpose, difficulty level, and constructs assessed by form A. During year 2 of the study, form A was administered during sessions 8 and 66 of intervention, and form B was administered during sessions 30 and 46. Over the course of year 3 of the study, form A was administered during sessions 8 and 46 of the intervention and form B was administered during sessions 30 and 66.

Lewkowicz (1980) categorized various methods of assessing phonological awareness abilities, many of which are used in the SBPAA. Elicited test behaviors that can indicate levels of PA include tapping out the number of sounds in a word, changing the order of sounds in a word, and putting together and taking apart sounds in a word (Lewkowicz, 1980). Four phonological tasks considered to be significant predictors of reading are discriminating rhymes, segmenting syllables, phoneme isolation, and phoneme blending (Bryant et al., 1990; Goswami & Bryant, 1990). Specifically, the SBPAA assesses PA skills by having students participate in a

variety of word games (Lane & Pullen, 2004). Tasks include (a) tapping out the number of words in a sentence, (b) deleting parts of words, (c) putting together parts of a word, (d) tapping syllables of a word, (e) deleting syllables from a word, (f) indicating whether words do or do not rhyme, (g) blending together the beginning of words, (h) generating their own rhymes, (i) blending together phoneme level parts of words, and (j) segmenting words at the phoneme level. See appendices A and B for forms A and B of the SBPAA. Furthermore, readers are referred to Table 2 for further elaboration on the composition and differences of forms A and B.

Table 2

Items and Tasks of Forms A and B of the SBPAA

	Total items	Word level	Syllable level	Onset Rime level	Phoneme level
Form A	60	10	15	15	20
Form B	55	10	15	10	20

During the PALS study, additional language and emergent literacy constructs were assessed with several pre-test and post-test measures and these results were used in an analysis of the convergent and discriminant validity of the SBPAA. Oral language skills, such as vocabulary and listening comprehension, were assessed using a variety of measures. These measures included the *Expressive One-Word Picture Vocabulary Test-IV* (EOWPVT-IV; Martin & Brownell, 2011) as a measure of expressive vocabulary. The *Peabody Picture Vocabulary Test-IV* (PPVT-IV; Dunn & Dunn, 2007) was used to measure receptive vocabulary. The Understanding Directions subtest from the *Woodcock-Johnson III Tests of Achievement* (WJ-III; Woodcock, McGrew, & Mather, 2001) was used to measure listening comprehension. Measured emergent literacy skills were phonological awareness, print knowledge, and reading ability. Phonological awareness was measured using the PA subtest of the *Test of Preschool Early Literacy* (TOPEL; Lonigan, Wagner, & Torgesen, 2007). Print knowledge was measured using

the Print Knowledge subtest of the TOPEL. Finally, word reading abilities were measured using the Letter-Word Identification subtest of the WJ-III.

Intervention

Participants in the PA treatment condition of the PALS study received individual tutoring for 15 to 20 minutes a session, four times a week for 20 weeks. The first 30 sessions of PA intervention activities were designed by the PI and graduate assistant. The second 30 lessons were take verbatim from *Road to the Code* (Blachman, 2000).

Procedures

Form equivalency. To address research question one, the data analytic plan was to average total scores from administrations 1 and 4 (form A) and compare them to the average of administrations 2 and 3 (form B). If growth is linear, this would provide adequate justification that the forms are parallel/equivalent. Allen and Yen (2002) speak to this process stating "a correlation between observed scores on alternate forms will produce a good estimate of test reliability if the alternate forms are parallel or if they are linear functions of parallel test scores and if carry-over effects and changes in scores over time do not influence the correlation" (p. 78).

Internal consistency reliability. To address research question two, an analysis of the internal consistency reliability of each administration of the SBPAA was conducted. Cronbach's alpha is one of the most commonly used measures of internal consistency in the social sciences (Allen & Yen, 2002; Meyer, 2010). Cronbach's alpha is symbolized as α and the formula is:

$$\text{Cronbach's } \alpha = \left(\frac{k}{k-1}\right)\left(\frac{\sigma_X^2 - \sum_{j=1}^{k} \sigma_{X_j}^2}{\sigma_X^2}\right)$$

A commonly used interpretation of Cronbach's alpha (Leech, Barrett, & Morgan, 2008) is as follows:

Excellent	Good	Acceptable	Questionable	Poor	Unacceptable
$\alpha \geq 0.9$	$0.9 > \alpha \geq 0.8$	$0.8 > \alpha \geq 0.7$	$0.7 > \alpha \geq 0.6$	$0.6 > \alpha \geq 0.5$	$0.5 > \alpha$

Item analysis. In order to address research question three, an item analysis of the SBPAA was conducted. Methods that were used included calculation of the difficulty and discrimination indices.

Difficulty index. The difficulty index of a measure is an item's difficulty level for the sample of examinees tested (Allen & Yen, 2002). In other words, for the set of examined individuals, it indicates how easy or difficult the items were. It is also referred to as the pass rate. Difficulty is indicated as p_1, and the formula is:

$$p_1 = \frac{\text{\# of correct answers}}{\text{\# of all answers}}.$$

Commonly used interpretation of the difficulty index is as follows:

	Very difficult	Fairly difficult	Moderately easy	Very easy
p_1	.00-.49	.50-.69	.70-.89	.90-1.00

The plan for analysis will include both the difficulty index at the item level and for each administration of the measure, i.e. time 1, time 2, etc.

Discrimination analysis. The discrimination index of a measure is an examination of the difference between the difficulty of a test item for upper and lower scoring subgroups of a population of examinees. In other words, it indicates if a test item is able to provide differentiation between high and low scoring individuals. The discrimination index is symbolized as D, and the formula is:

$$D = p_{\text{upper group}} - p_{\text{lower group}}$$

Commonly used interpretation of the discrimination index is as follows:

Satisfactory	Little/no revision	Revise	Complete Revision	drop

D_i	>=.40	.30-.39	.20-.29 (.16-.29)	<=.19 (in reality .15)	0<.05

The plan for analysis included the discrimination index for item level data across each administration of the SBPAA.

Convergent validity. In order to address research question four, an analysis of the convergent validity of the SBPAA was conducted. Convergent validity is determined by exploring correlations between scores that measure the same trait via different tests/methods (Allen & Yen, 2002). The convergent validity of the SBPAA was assessed by using regression analysis to compare the results of its final administration to the results from PA sections of the normed and validated *Test of Preschool Early Literacy*-TOPEL (Lonigan et al., 2007) administered several weeks later. Crocker and Algina (1986) maintain that a correlation coefficient of 0.80 to 0.90 on similar individually administered measures demonstrates strong similarity between an individual's performance on two different tests.

Ability to detect change. In order to address question five, paired t-tests were used to examine if there were statistically significant differences in obtained scores between sequential administrations of forms A and B of the SBPAA.

Chapter IV: Results

Form Equivalency

Research Question 1: *What level of equivalency* do *form A and the alternate form B of the Sound Beginning Phonological Awareness Assessment (SBPAA) possess?* Research Question 1, cannot be answered at this time. The possibility of answering this question was based on the presumption of a linear growth rate for the subjects' phonological awareness skills during intervention. Upon analysis of student growth in the PA condition it was discovered that the growth rate followed a non-linear trajectory. In order to compare the two test administrations, the SBPAA form A had 5 items removed in order to have it match form B and have an equal number of matching items between the two forms. Despite this change which attempted to make the two forms more equivalent, the four administrations did not follow a linear growth trajectory. For further illustration, readers are referred to Figure 4. Therefore, assumptions required to equate the two forms cannot be met at this time.

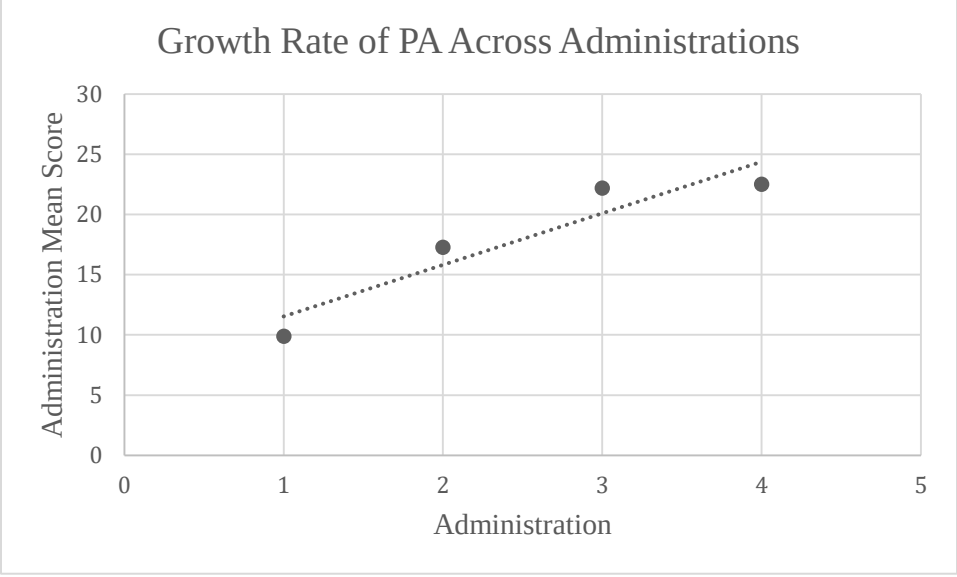

Figure 4. Mean Score of Administrations of the SBPAA.

Readers are referred to appendix C for further reference on the order of administration of forms A and B to participants. Solutions to these methodological issues will be addressed further in the limitations and future directions section.

Reliability

Research Question 2: *What is the measure of internal consistency for the SBPAA when used with preschoolers with ASD?* Cronbach's alpha was computed for each of the four administrations of the SBPAA. Readers are directed to Table 2 for these results. Both form A and form B were used with the 2013 and 2014 PA intervention cohorts. However, the order of administration varied between years. Readers are directed to Figure 5 for further information regarding which form of the SBPAA was administered during which lesson and administration.

	Admin 1	Admin 2	Admin 3	Admin 4
2013 cohort	A	B	B	A
2014 cohort	A	B	A	B
Lesson Administered	Lesson 8	Lesson 30	Lesson 46	Lesson 66

Figure 5. Order of administration of SBPAA forms A and B across intervention cohorts.

Reliability data from all administrations of the SBPAA are represented in Table 3. Overall, both forms of the SBPAA had Cronbach's alpha measures ranging from 0.95 – 0.98. Administration 1 of form A (n = 40) was composed of 60 items (M = 14.97, SD = 12.01) and produced a Cronbach's Alpha score of 0.95. Administration 2 of form B (n = 34) was composed of 55 items (M = 23.26, SD = 15.40) and produced a Cronbach's Alpha score of 0.97. Administration 3 of form A (n = 11) was composed of 60 items (M = 33.00, SD = 15.93) and produced a Cronbach's Alpha score of 0.96. Administration 3 of form B (n = 24) was composed of 55 items (M = 23.21, SD = 16.69) and produced a Cronbach's Alpha score of 0.97.

Administration 4 of form A (n = 16) was composed of 60 items (M = 24.38, SD = 19.02) and produced a Cronbach's Alpha score of 0.98. Administration 4 of form B (n = 4) was composed of 55 items (M = 36.25, SD = 15.99) and produced a Cronbach's Alpha score of 0.97.

Table 3

Cronbach's Alpha Scores of SBPAA forms A and B Across Administrations

Administration	*n*	Cronbach's Alpha	Number of items	Mean	Variance	Std. Deviation
Admin 1 (form A) 2013 & 2014 cohorts	40	0.95	60	14.97	133.38	12.02
Admin 2 (form B) 2013 & 2014 cohorts	34	0.97	55	23.26	236.99	15.40
Admin 3 (form B) 2013 cohort	24	0.97	55	23.21	278.52	16.69
Admin 3 (form A) 2014 cohort	11	0.96	60	33.00	253.80	15.93
Admin 4 (form A) 2013 cohort	16	0.98	60	24.38	361.58	19.02
Admin 4 (form B) 2014 cohort	4	0.97	55	36.25	255.58	15.99

Difficulty, Discrimination, and Discriminability

Research Question 3: *What are the levels of item difficulty, discrimination, and discriminability for forms A and B of the SBPAA?* Examination of item level data for the administration of each form of the SBPAA with the highest number of participants was conducted in order to obtain sufficient variability in regards to item level performance. Question three included two subparts which were: *(1)* an analysis of the difficulty index of each item; and (2) the ability of each item to discriminate between upper and lower performing subgroups of participants. Analyses are grouped into word level, syllable level, onset-rime level, and phoneme level groupings. For each administration and assessment level, the difficulty index (passing rate), discrimination, and discriminability were calculated.

Word level form A. Word-level items of administration 1 of form A (n = 40) were administered to participants. Readers should refer to Table 4 for full analysis. The word-level section consisted of 10 items (M = 3.30, SD = 3.20) and produced a Cronbach's Alpha score of 0.88, which is within the Good range for internal consistency. All the items had means within the Very Difficult or Fairly Difficult range. Eight of the 10 items had Satisfactory levels of discriminability. Item five may require little to no revision and item three may require complete revision. The alpha level remains the same or goes down if any item is deleted, except for Q1, which indicates those items significantly contribute to the value of the subscale. If Q1 is deleted the alpha level goes up, however only by 0.01, which does not meet the threshold for dropping or revising the item. The corrected item-total correlations are all positive and relatively high.

Table 4

Word Level Item Data from Administration 1 of Form A of the SBPAA

Number of Items	Cronbach's Alpha	Mean	Variance	Std. Deviation
10	0.88	3.30	10.27	3.20

Item	Item Means	Upper Mean	Lower Mean	Discrim.	Corrected Item-Total Correlation	Alpha if Item Deleted
Q1	0.50	0.85	0.23	0.62	0.37	0.89
Q2	0.50	0.77	0.15	0.62	0.62	0.87
Q3	0.45	0.31	0.15	0.16	0.39	0.88
Q4	0.51	0.77	0.15	0.62	0.56	0.87
Q5	0.45	0.38	0.08	0.30	0.51	0.88
Q6	0.45	0.62	0	0.62	0.73	0.86
Q7	0.46	0.69	0	0.69	0.80	0.85
Q8	0.42	0.62	0	0.62	0.74	0.86
Q9	0.42	0.62	0	0.62	0.72	0.86
Q10	0.44	0.62	0	0.62	0.70	0.86

Syllable level form A. Syllable-level items of administration 1 of form A (n = 40) were administered to participants. Readers should refer to Table 5 for full analysis. The syllable-level

section consisted of 15 items (*M* = 5.33, *SD* = 3.66) and produced a Cronbach's Alpha score of 0.83, which is within the Good range for internal consistency. All the items had means within the Very Difficult or Fairly Difficult range. Ten of the 15 items had Satisfactory levels of discriminability. Items 17, 22, 24, and 25 may require little to no revision and item 15 should be revised. The alpha level remains the same or goes down if any item is deleted indicating that all items significantly contribute to the value of the subscale. The corrected item-total correlations are all positive and relatively high, excluding item 17 with a 0.25 correlation.

Table 5

Syllable Level Item Data from Administration 1 of Form A of the SBPAA

Number of Items	Cronbach's Alpha	Mean	Variance	Std. Deviation
15	0.83	5.33	13.40	3.66

Item	Item Means	Upper Mean	Lower Mean	Discrim.	Corrected Item-Total Correlation	Alpha if Item Deleted
Q11	0.53	0.69	0.23	0.46	0.36	0.82
Q12	0.50	0.77	0.15	0.62	0.49	0.82
Q13	0.35	0.77	0.08	0.69	0.71	0.80
Q14	0.35	0.69	0.08	0.61	0.52	0.81
Q15	0.10	0.23	0	0.23	0.43	0.82
Q16	0.43	0.69	0	0.69	0.40	0.82
Q17	0.55	0.62	0.23	0.39	0.25	0.83
Q18	0.60	0.92	0.15	0.77	0.65	0.80
Q19	0.53	0.69	0.23	0.46	0.39	0.82
Q20	0.55	0.69	0.23	0.46	0.35	0.82
Q21	0.18	0.46	0	0.46	0.54	0.81
Q22	0.13	0.31	0	0.31	0.37	0.82
Q23	0.25	0.62	0.08	0.54	0.57	0.81
Q24	0.15	0.38	0.08	0.30	0.34	0.82
Q25	0.15	0.38	0.08	0.30	0.43	0.82

Onset rime level form A. Onset-rime-level items of administration 1 of form A (*n* = 40) were administered to participants. Readers should refer to Table 6 for full analysis. The syllable-level section consisted of 15 items (*M* = 4.08, *SD* = 3.85) and produced a Cronbach's

Alpha score of 0.87, which is within the Good range for internal consistency. All the items had means within the Very Difficult range, excluding items 26 and 29, which were in the Fairly Difficult range. Four of the 15 items had Satisfactory levels of discriminability. Items 27, 28, 29, 32, 36, 37, 39, and 40 may require little to no revision, and items 30 and 38 should be revised. Item 34 falls within the range of discriminability that calls for complete revision. The alpha level remains the same or goes down for all items except 27, 29, and 30. However, if these items are deleted the alpha level goes up by only 0.01, which does not meet the threshold for dropping or revising the item. The corrected item-total correlations are all positive and relatively high, excluding item 27 with a 0.17 correlation.

Table 6

Onset Rime Level Item Data from Administration 1 of Form A of the SBPAA

Number of Items	Cronbach's Alpha	Mean	Variance	Std. Deviation
15	0.87	4.08	14.84	3.85

Item	Item Means	Upper Mean	Lower Mean	Discrim.	Corrected Item-Total Correlation	Alpha if Item Deleted
Q26	0.58	0.85	0.08	0.77	0.51	0.87
Q27	0.23	0.38	0	0.38	0.17	0.88
Q28	0.48	0.62	0.23	0.39	0.39	0.87
Q29	0.53	0.62	0.31	0.31	0.33	0.88
Q30	0.23	0.38	0.15	0.23	0.30	0.88
Q31	0.28	0.77	0.08	0.69	0.71	0.86
Q32	0.28	0.46	0.15	0.31	0.59	0.86
Q33	0.35	0.69	0.23	0.67	0.56	0.86
Q34	0.20	0.38	0.23	0.15	0.49	0.87
Q35	0.28	0.62	0.15	0.47	0.59	0.86
Q36	0.18	0.38	0	0.38	0.70	0.86
Q37	0.13	0.38	0	0.38	0.77	0.86
Q38	0.10	0.23	0	0.23	0.63	0.86
Q39	0.13	0.31	0	0.31	0.68	0.86
Q40	0.15	0.38	0	0.38	0.79	0.86

Phoneme level form A. Phoneme-level items of administration 1 of form A (n = 40) were administered to participants. Readers should refer to Table 7 for full analysis. No participants correctly answered items 58 and 60, therefore item level data on these items was not available. The phoneme-level section consisted of 20 items (M = 2.28, SD = 3.69) and produced a Cronbach's Alpha score of 0.91, which is within the Excellent range for internal consistency. All the items had means within the Very Difficult range. Only item 43 had Satisfactory levels of discriminability. Items 42, 44, 45, 47, 48, and 52 may require little to no revision. Items 46, 49, 53, 55, and 56 should be revised. Items 41, 50, 51, 54, 57, and 59 fall within the range of discriminability that calls for complete revision. Items 58 and 60 should be dropped. The alpha

level remains the same or goes down for all items within this section. The corrected item-total correlations are all positive and relatively high.

Table 7

Phoneme Level Item Data from Administration 1 of Form A of the SBPAA

Number of Items	Cronbach's Alpha	Mean	Variance	Std. Deviation
20	0.91	2.28	13.64	3.69

Item	Item Means	Upper Mean	Lower Mean	Discrim.	Corrected Item-Total Correlation	Alpha if Item Deleted
Q41	0.44	0.31	0.15	0.16	0.41	0.91
Q42	0.45	0.46	0.08	0.38	0.53	0.91
Q43	0.45	0.62	0.15	0.47	0.60	0.90
Q44	0.36	0.38	0	0.38	0.71	0.90
Q45	0.41	0.46	0.08	0.38	0.80	0.90
Q46	0.27	0.23	0	0.23	0.58	0.90
Q47	0.30	0.31	0	0.31	0.79	0.90
Q48	0.39	0.46	0.08	0.38	0.84	0.90
Q49	0.34	0.31	0.08	0.23	0.72	0.90
Q50	0.22	0.15	0	0.15	0.41	0.91
Q51	0.27	0.08	0	0.08	0.30	0.91
Q52	0.36	0.38	0	0.38	0.47	0.91
Q53	0.27	0.23	0	0.23	0.50	0.91
Q54	0.22	0.15	0	0.15	0.47	0.91
Q55	0.30	0.23	0	0.23	0.46	0.91
Q56	0.27	0.23	0	0.23	0.73	0.90
Q57	0.22	0.15	0	0.15	0.64	0.90
Q58	--	--	--	--	--	--
Q59	0.16	0.08	0	0.08	0.62	0.91
Q60	--	--	--	--	--	--

Word level form B. Word-level items of administration 2 of form B (n = 34) were administered to participants. Readers should refer to Table 8 for full analysis. The word-level section consisted of 10 items (M = 4.15, SD = 3.50) and produced a Cronbach's Alpha score of 0.90, which is within the Excellent range for internal consistency. Items 1 and 4 had means within the Fairly Difficult range, while the rest of the items produced means within the Very

Difficult range. All 10 items had Satisfactory levels of discriminability. The alpha level goes down if any item is deleted, except for item 2, which indicates all other items significantly contribute to the value of the subscale. If item 2 is deleted the alpha level goes up, however only by 0.01, which does not meet the threshold for dropping or revising the item. The corrected item-total correlations are all positive and relatively high.

Table 8

Word Level Item Data from Administration 2 of Form B of the SBPAA

Number of Items	Cronbach's Alpha	Mean	Variance	Std. Deviation
10	0.90	4.15	12.25	3.50

Item	Item Means	Upper Mean	Lower Mean	Discrim.	Corrected Item-Total Correlation	Alpha if Item Deleted
Q1	0.68	0.91	0.18	0.73	0.52	0.90
Q2	0.41	0.64	0.18	0.46	0.36	0.91
Q3	0.26	0.73	0	0.73	0.59	0.89
Q4	0.59	0.91	0.18	0.73	0.51	0.90
Q5	0.29	0.64	0	0.64	0.60	0.89
Q6	0.38	0.82	0	0.82	0.77	0.88
Q7	0.41	0.82	0	0.82	0.78	0.88
Q8	0.35	0.82	0	0.82	0.80	0.88
Q9	0.32	0.82	0	0.82	0.80	0.88
Q10	0.44	0.91	0	0.91	0.76	0.88

Syllable level form B. Syllable-level items of administration 2 of form B (n = 34) were administered to participants. Readers should refer to Table 9 for full analysis. The syllable-level section consisted of 15 items (M = 8.74, SD = 4.50) and produced a Cronbach's Alpha score of 0.90, which is within the Excellent range for internal consistency. Syllable-level items on this section ranged from Moderately Easy to Fairly Difficult. Items 11, 12, 16, 17, and 20 produced means within the Moderately Easy range. Items 13, 14, 15, 18, and 19 produced means within the Fairly Difficult range. Items 21 though 25 produced means within the Very Difficult range. Thirteen of the 15 items had Satisfactory levels of discriminability. Item 20 may require little to

no revision and item 17 should be revised. The alpha level remains the same or goes down if any item is deleted indicating that all items significantly contribute to the value of the subscale. The corrected item-total correlations are all positive and relatively high.

Table 9

Syllable Level Item Data from Administration 2 of Form B of the SBPAA

Number of Items	Cronbach's Alpha	Mean	Variance	Std. Deviation
15	0.90	8.74	20.26	4.50

Item	Item Means	Upper Mean	Lower Mean	Discrim.	Corrected Item-Total Correlation	Alpha if Item Deleted
Q11	0.71	0.91	0.36	0.55	0.51	0.90
Q12	0.71	0.91	0.45	0.46	0.48	0.90
Q13	0.65	0.91	0.27	0.64	0.58	0.89
Q14	0.68	1.0	0.18	0.82	0.75	0.89
Q15	0.50	0.82	0.09	0.73	0.65	0.89
Q16	0.71	0.91	0.36	0.55	0.55	0.90
Q17	0.82	0.91	0.64	0.27	0.43	0.90
Q18	0.68	0.91	0.27	0.64	0.49	0.90
Q19	0.65	0.73	0.36	0.46	0.39	0.90
Q20	0.88	1.0	0.64	0.36	0.55	0.90
Q21	0.47	0.91	0	0.91	0.68	0.89
Q22	0.38	0.91	0	0.91	0.69	0.89
Q23	0.32	0.91	0	0.91	0.68	0.89
Q24	0.29	0.64	0	0.64	0.62	0.89
Q25	0.29	0.73	0	0.73	0.69	0.89

Onset rime level form B. Onset-rime-level items of administration 2 of form B (n = 34) were administered to participants. Readers should refer to Table 10 for full analysis. The onset rime level section consisted of 10 items (M = 5.24, SD = 3.57) and produced a Cronbach's Alpha score of 0.90, which is within the Excellent range for internal consistency. Items 26, 28, and 29 fell within the Moderately Easy range. Items 31 and 32 fell within the Fairly Difficult range, while items 27, 30, 33, 34, and 35 fell within the Very Difficult range. Nine of the 10 items had Satisfactory levels of discriminability. Item 30 may require revision. The alpha level

goes down if any item, except for item 30, is deleted, which indicates all other items significantly contribute to the value of the subscale. If item 30 is deleted the alpha level goes up. However it increases by only by 0.01, which does not meet the threshold for dropping or revising the item. The corrected item-total correlations are all positive and relatively high.

Table 10

Onset Rime Level Item Data from Administration 2 of Form B of the SBPAA

Number of Items	Cronbach's Alpha	Mean	Variance	Std. Deviation
10	0.90	5.24	12.73	3.57

Item	Item Means	Upper Mean	Lower Mean	Discrim.	Corrected Item-Total Correlation	Alpha if Item Deleted
Q26	0.74	1.0	0.36	0.64	0.65	0.90
Q27	0.32	0.64	0.09	0.55	0.52	0.90
Q28	0.71	1.0	0.27	0.73	0.63	0.90
Q29	0.71	1.0	0.27	0.73	0.63	0.90
Q30	0.38	0.55	0.27	0.28	0.35	0.91
Q31	0.53	1.0	0	1.0	0.85	0.88
Q32	0.53	1.0	0	1.0	0.78	0.89
Q33	0.47	0.91	0	0.91	0.73	0.89
Q34	0.44	0.82	0	0.82	0.70	0.89
Q35	0.41	1.0	0	1.0	0.78	0.89

Phoneme level form B. Phoneme-level-items of administration 2 of form B (n = 34) were administered to participants. Readers should refer to Table 11 for full analysis. The phoneme level section consisted of 20 items (M = 5.15, SD = 5.47) and produced a Cronbach's Alpha score of 0.93, which is within the Excellent range for internal consistency. All but two of the 20 items fell within the Very Difficult range. Items 36 and 38 fell within the Fairly Difficult range. Eleven of the 20 items had Satisfactory levels of discriminability. Items 47 and 49 may require little to no revision. Items 50, 52, and 54, may require revision. Items 51, 53, and 55 should be completely revised. The alpha level goes down if any item is deleted, which indicates all items

significantly contribute to the value of the subscale. The corrected item-total correlations are all positive and relatively high.

Table 11

Phoneme Level Item Data from Administration 2 of Form B of the SBPAA

Number of Items	Cronbach's Alpha	Mean	Variance	Std. Deviation
20	0.93	5.15	29.95	5.47

Item	Item Means	Upper Mean	Lower Mean	Discrim.	Corrected Item-Total Correlation	Alpha if Item Deleted
Q36	0.59	1.0	0.09	0.91	0.69	0.93
Q37	0.44	1.0	0	1.0	0.81	0.92
Q38	0.50	0.91	0.09	0.82	0.71	0.93
Q39	0.35	0.82	0	0.82	0.70	0.93
Q40	0.44	0.91	0.09	0.82	0.76	0.92
Q41	0.32	0.82	0	0.82	0.76	0.92
Q42	0.26	0.73	0	0.73	0.70	0.93
Q43	0.29	0.82	0	0.82	0.78	0.92
Q44	0.18	0.45	0	0.45	0.60	0.93
Q45	0.18	0.55	0	0.55	0.71	0.93
Q46	0.35	0.73	0	0.73	0.69	0.93
Q47	0.21	0.45	0.09	0.36	0.54	0.93
Q48	0.26	0.45	0	0.45	0.55	0.93
Q49	0.18	0.36	0	0.36	0.51	0.93
Q50	0.15	0.27	0	0.27	0.46	0.93
Q51	0.12	0.18	0	0.18	0.38	0.93
Q52	0.12	0.27	0	0.27	0.52	0.93
Q53	0.06	0.18	0	0.18	0.45	0.93
Q54	0.12	0.27	0	0.27	0.34	0.93
Q55	0.03	0.09	0	0.09	0.39	0.93

Convergent Validity

Research Question 4: What is the degree of convergent validity between the SBPAA and the TOPEL on a sample of preschool children with ASD? Examination of the convergent and discriminant validity between the SBPAA and various pre and post-test measures was estimated by using correlation coefficients. Readers are directed to Tables 12 through 15 for these results.

Form A administration 1 and form B administration 2 were compared to pre-test measures as these administrations of the SBPAA occurred during the first half of intervention. Form A admin 4 and form B administration 3 were compared to post-test measures as these administrations of the SBPAA occurred during the second half of intervention.

Correlation statistics for a comparison between form A administration one of the SBPAA and pre-test measures are displayed below in Table 12. By examining this table one can see that the SBPAA was highly correlated with the TOPEL ($r = 0.69$, $p < 0.001$). The scores from this administration of the SBPAA were also correlated with the *Test of Print Knowledge* ($r = 0.63$, $p < 0.001$). The scores from this administration were less correlated with oral language measures such as the two vocabulary tests (EOWPVT, PPPVT) and the WJ Understanding Directions, indicating degrees of discriminant validity.

Table 12

Correlations Between SBPAA Form A and Pre-Test Measures

Measure	1.	2.	3.	4.	5.	6.
1. Form A Administration 1	--					
2. EOWPVT	.49 **	--				
3. PPVT	.45 **	.66 ***	--			
4. Test of Print Knowledge	.63 ***	.47 **	.52 ***	--		
5. Test of Phonological Awareness	.69 ***	.51 ***	.78 ***	.69 ***	--	
6. WJ Understanding Directions	.49 **	.72 ***	.77 ***	.49 **	.54 ***	--

Note. N=37. * $p < .05$, ** $p < .01$, *** $p < .001$. Measure 2 – 6 are pre-test measures.

Correlation statistics for a comparison between form A administration 4 of the SBPAA and post-test measures are displayed in Table 13. By examining this table one can see that the SBPAA was highly correlated with the TOPEL ($r = 0.86$, $p < 0.001$). The scores from this administration of the SBPAA were also correlated with the *Test of Print Knowledge* ($r = 0.84$, $p < 0.001$). The scores from this administration were less correlated with oral language measures

such as the EOWPVT, PPVT, and WJ Understanding Directions, indicating degrees of discriminant validity.

Table 13

Correlations Between SBPAA Form A and Post-Test Measures

Measure	1.	2.	3.	4.	5.	6.
1. Form A Administration 4	--					
2. EOWPVT	.64 **	--				
3. PPVT	.51 *	.89 ***	--			
4. Test of Print Knowledge	.84 ***	.58 ***	.56 ***	--		
5. Test of Phonological Awareness	.86 ***	.80 ***	.76 ***	.73 ***	--	
6. WJ Understanding Directions	.63 *	.75 ***	.77 ***	.60 ***	.74 ***	--

Note. N=15. * $p < .05$, ** $p < .01$, *** $p < .001$. Measure 2 – 6 are post-test measures.

Correlation statistics for a comparison between form B administration 2 of the SBPAA and pre-test measures are displayed in Table 14. By examining this table one can see that the SBPAA approaches high levels of correlation with the TOPEL ($r = 0.77$, $p < 0.001$). The scores from this administration of the SBPAA were also correlated with the *Test of Print Knowledge* ($r = 0.68$, $p < 0.001$). The scores from this administration were less correlated with oral language measures such as the EOWPVT, PPVT, and WJ Understanding Directions, indicating degrees of discriminant validity.

Table 14

Correlations Between SBPAA Form B and Pre-Test Measures

Measure	1.	2.	3.	4.	5.	6.
1. Form B Administration 2	--					
2. EOWPVT	.56 ***	--				
3. PPVT	.57 ***	.66 ***	--			
4. Test of Print Knowledge	.68 ***	.47 **	.52 ***	--		
5. Test of Phonological Awareness	.77 ***	.51 ***	.78 ***	.65 ***	--	
6. WJ Understanding Directions	.51 **	.72 ***	.77 ***	.49 **	.54 ***	--

Note. N=32. * p < .05, ** p < .01, *** p < .001. Measure 2 – 6 are pre-test measures.

Correlation statistics for a comparison between form B administration 3 of the SBPAA and post-test measures are displayed below in Table 15. By examining this table one can see that the SBPAA was highly correlated with the TOPEL ($r = 0.89$, $p < 0.001$). The scores from this administration of the SBPAA were also moderately correlated with the *Test of Print Knowledge* ($r = 0.77$, $p < 0.001$). The scores from this administration were less correlated with oral language measures such as the EOWPVT, PPPVT, and WJ Understanding Directions, indicating degrees of discriminant validity.

Table 15

Correlations Between SBPAA Form B and Post-Test Measures

Measure	1.	2.	3.	4.	5.	6.
1. Form B Administration 3	--					
2. EOWPVT	.74 ***	--				
3. PPVT	.72 ***	.89 ***	--			
4. Test of Print Knowledge	.77 ***	.58 ***	.56 ***	--		
5. Test of Phonological Awareness	.89 ***	.80 ***	.76 ***	.73 ***	--	
6. WJ Understanding Directions	.69 ***	.75 ***	.77 ***	.59 ***	.74 ***	--

Note. N=22. * p < .05, ** p < .01, *** p < .001. Measure 2 – 6 are post-test measures.

Ability to Detect Change

Research Question 5: To what degree is the SBPAA able to measure growth of phonological awareness skills in preschoolers with ASD? In order to address Question 5, paired t-tests were used to examine if there were statistically significant differences in obtained scores between sequential administrations of forms A and B of the SBPAA. Readers are directed to Table 16 for these results.

Paired t-tests revealed that there were statistically significant differences between sequential measures of both form A and form B of the SBPAA. Examination of the results of a paired sample t-test that compared administration 4 of form A (M = 20.86, SD = 16.24) and administration 1 of form A (M = 9.47, SD = 9.60) showed statistically significant differences $t(14) = 5.07$, $p < .001$ and an effect size of $d = 1.31$. Furthermore, examination of the results of a paired sample t-test that compared administration 3 of form B (M = 23.60, SD = 16.97) and administration 2 of form B (M = 19.05, SD = 15.75) showed statistically significant differences $t(19) = 4.46$, $p < .001$ and an effect size of $d = 0.99$.

Table 16

Paired Sample t-Tests

Measures	*M*	*(SD)*	*n*	*t*	*p*
SBPAA Form A			15	*t*(14)	
Administration 4	20.86	(16.24)			
Administration 1	9.47	(9.60)			
Paired Sample	11.40	(8.72)		5.07	<.001
SBPAA Form B			20	*t*(19)	
Administration 3	23.60	(16.97)			
Administration 2	19.05	(15.75)			
Paired Sample	4.55	(4.56)		4.46	<.001

Chapter V: Discussion

Summary and Implications of Findings

Overall, the purpose of this study was to evaluate the psychometric properties of a Phonological Awareness (PA) progress monitoring tool for use with preschool students on the Autism spectrum. This study had several goals. First, it sought evidence regarding whether an alternate form of the SBPAA had sufficient levels of equivalency. Second, it sought to determine if sufficient levels of reliability were present when used with this population of individuals. Third, an item level analysis was conducted to determine how items performed when used with preschoolers with ASD. Fourth, information regarding concurrent and discriminant validity with other measures of phonological awareness and early literacy skills was analyzed. Finally, an analysis was conducted to verify if the progress monitoring measure has sufficient levels of sensitivity to detect changes in PA skills. Further implications of these results will be discussed, along with limitations to the study and suggested future directions for research.

Equivalence

Because the growth model was not linear, the current data do not support that form A and alternate form B are equivalent. This does not mean that the two do not have equivalency, it simply means that that conclusion cannot be supported with the current data set. Therefore, the two tests were treated like separate measures during analysis of the subsequent four questions. To address the equivalency question, further research will need to be conducted and likely include larger sample sizes to enhance power to examine this question.

Reliability

An analysis of these administrations of the SBPAA to preschoolers with ASD demonstrated good overall levels of internal consistency reliability. The fields of education,

psychology, and measurement have established guidelines for acceptable levels of reliability for assessment measures (AERA, APA, & NCME, 2014). Internal consistency reliability was assessed using Cronbach's alpha, one of the most commonly used measures of internal consistency in the social sciences (Allen & Yen, 2002; Meyer, 2010). All administrations of both form A and form B of the SBPAA produced Cronbach's Alpha scores ranging from 0.95 to 0.98. These levels of reliability fall above the minimum acceptable levels required for excellent validity, as established by Leech, Barrett, and Morgan (2008) and by George and Mallery (2003). This means that both forms of the SBPAA can be used with preschoolers with ASD while adhering to the high standards of psychometrics.

Notable were the high levels of variance found during the analysis. The high level of variance is likely due to the range of skill levels exhibited by this sample of preschoolers with ASD. Their performance on different administrations often produced a somewhat platykurtic distribution of test scores. Despite, this the SBPAA appeared to produce consistently reliable results.

Item Analysis

Due to a need for a sufficient number of participants analyses were conducted on administrations 1 and 2 of forms A and B. When this study was initially proposed, an item level analysis for administrations 3 and 4 was also planned, but due to a lack of sufficient data this was not possible. The first two administrations had the highest collective numbers of participants which allowed for the necessary levels of variability for valid analysis. As previously mentioned, a somewhat platykurtic curve resulted from the scores. This occurs when a test is administered with items of moderate difficulty and good discriminating power (Allen & Yen, 2002).

Word level form A. Scores from the administration of this subtest indicate that overall the subtest had good levels of reliability and had many Very Difficult items, with just less than half of students answering items correctly. Generally, item level difficulty is recommended to be between 0.3 – 0.7 (Allen & Yen, 2002). However, because this is a diagnostic measure and progress monitoring tool, this emphasis was lowered when evaluating item adequacy. Most items showed high levels of discrimination, which indicated good overall differentiation between high and low performers. All items met Pallant's (2007) suggested corrected item-total correlation cut score of 0.3 for inclusion. Only one question, item Q1, approached this level, and the Cronbach's alpha level of the subtest would go up if the item was deleted. This means that whether a student answered this item correctly was not highly predictive of their overall score on the entire measure. As this is a relatively difficult item ($p = 0.50$) revision may be warranted for use with this population. It is often best to begin tests with easy items and move toward more difficult items. The item response data indicated that half of all participants incorrectly answer this first item. Thus it may be warranted to revise the item so that more students correctly answer the very first question they are exposed to on the test.

Syllable level form A. Scores from the administration of this subtest indicate that overall, the subtest had good levels of reliability and had many Very Difficult items, with just less than half of students answering items correctly. Floor effects may have an impact on this subtest, as items Q15, Q22, Q24, and Q25 produced item means ≤ 0.15. However, revision of these items should be approached with caution, as in order to detect growth, especially in strong performers, some item difficulties must be relatively high as to avoid the alternative problem of ceiling effects during later stages of intervention. Revision of item Q17 should be considered, as it

produced a corrected item-total correlation below the suggested cut score of 0.3 and thus may not significantly contribute to the value of the subscale (Allen and Yen, 2002; Pallant, 2007).

Onset rime level form A. All the items had means in the Very Difficult range, except items Q26 and Q29, both of which had mean scores that fell within the Fairly Difficult range. Only four of the 15 items had difficulty ratings within the range suggested by Allen and Yen (2002). However, later administration of form B demonstrated significant growth on the part of participants. None of the difficulty levels fell below 0.3 on that administration when focusing on onset-rime level skills. Therefore, it is advised to retain these items in order to provide a range of items capable of detecting skill acquisition. Items Q27, Q29, and Q30 were close to or below Pallant's (2007) suggested corrected item total correlation cut-score of 0.3. The alpha level of this subtest also increased if these items are deleted, indicating a possible need for revision. Again, similar items designed to measure the skill on form B had difficulty levels and corrected item-total correlations within acceptable limits, suggesting that as a progress monitoring tool these items significantly contribute to the test, and although some items may need revision, they should not be dropped. Because children, especially preschool aged children with ASD, do not have an excess of sustained attention or executive functioning skills (Liss et al., 2001), it is essential to gather as much good assessment data as quickly as possible. If test items do not contribute to the overall test by improving differentiation between children who have or have not acquired a skill or helping to by helping to predict a child's overall skill level in the area, then the items should be replaced or revised. It is important to have technically adequate assessment items because it is essential to ensure time spent administering test items provides useful information for educators.

Phoneme level form A. On this administration, items Q41-Q60 produced unusual results. The subtest had exceedingly high levels of variance. It also had Very Difficult items and had less consistent discriminability than other subtests on form A. There was also great variability between items in regards to corrected item-total correlations. This suggested that these items, which are designed to measure phoneme level skills, do not consistently predict total test performance at this point during intervention. Curiously, several items with lower rates of corrected item-total correlation actually had item performances which resulted in item means exceeding the upper means. This indicated that the best performers in regards to total test score did not consistently perform better on this subset of phoneme level items than individuals who demonstrated average levels of performance in terms of total test score. The Cronbach's alpha score of the subtest indicates that it demonstrates high levels of internal consistency despite a large number of items with poor discriminability. This may have been due in part to the fact that this was the first administration of the SBPAA, which may have resulted in some floor effects as the subtest measured skills that were likely beyond the pre-intervention repertoire of preschoolers with ASD.

Word level form B. Scores from the administration of this subtest indicated that overall the subtest produced excellent levels of reliability and had many very difficult items; nine of the 10 questions produced difficulty scores within the Very Difficult range. As previously mentioned regarding form A, item level difficulty is recommended to be between 0.3 – 0.7 (Allen & Yen, 2002). However, because this is a diagnostic measure and a progress monitoring tool this emphasis is lowered when evaluating item adequacy; only item Q3 fell just outside this range ($p_1 = 0.26$). Like form A, most items showed high levels of discrimination, which indicates good overall differentiation between high and low performers. All items meet Pallant's

(2007) suggested corrected item-total correlation cut score of 0.3 for inclusion. However, item Q2 approaches this level, and the Cronbach's Alpha level of the subtest increases if the item is deleted. As this is a Very Difficult item (p = 0.41) revision may be warranted for use with this population. Revision of items lacking the highest levels of technical adequacy is important because as previously mentioned, efficiency when testing this population is essential. Researchers have documented that children with ASD often have a more difficult time with sustained attention, largely due to developmental delays and motivation (Garretspm. Fein, & Waterhouse, 1990).

Syllable level form B. Scores from the administration of this subtest indicate that overall, the subtest produced excellent levels of reliability and had more variability regarding difficulty level of items. Although the two forms cannot be equated at this time, syllable level results for forms A and B were very similar in that the upper third of participants performed at a level resulting in item means mostly in the Very Easy and Moderately Easy ranges. The lower group of performers also demonstrated overall growth in performance. However, despite this growth, the form B subtest still demonstrated high levels of discriminability with the exception of item Q17. However, unlike item Q17 on form A, item Q17 on form B produced a higher score for its corrected item-total correlation at 0.43 (Allen and Yen, 2002; Pallant, 2007). All 15 items appeared to significantly contribute to the syllable level subscale for form B. Finding syllable level skill growth in the lowest performing participants is promising. This finding suggests two things. First, the assessment measure is sensitive to skill growth, especially in low performers who would likely require intervention. Therefore the assessment would be useful in an educational setting for progress monitoring, such as within an RtI framework. Second, though

beyond the scope of the current study, skill growth in this area would imply that low performing preschoolers with ASD might benefit from phonological awareness intervention.

Onset Rime level form B. This level of subtest is more difficult to directly compare to its equivalent subtest on form A. Unlike form A which had 15 items, form B had only 10 items at the Onset Rime level. While both form A and form B included items which probed for skills related to matching rhymes and to generating rhymes, only form A had items probing for skills related to blending onsets and rimes. Form B did not include assess onset rime blending. At the Onset Rime level of form B, item difficulty levels ranged from Moderately Easy to Fairly Difficult. On form A most items' means were in the Fairly Difficult range, suggesting with time and intervention participants are showing growth on these skills. With the exception of item Q30, all items on form B showed satisfactory levels of discriminability. If item Q30 is deleted the alpha level of the subtest goes up, therefore some revision of this item may be needed.

Phoneme level form B. Item performance on this subtest was less erratic than its related subtest on administration 1 of form A. While the subtest continued to have high levels of variance, it did have two items with means in the Fairly Difficult range and the rest in the Very Difficult range. However, the top third of performers showed growth in their ability to blend phonemes together into words, as required in items Q36-Q45. This group performed relatively well on words of up to three phonemes. Words with four and five phonemes still proved to be quite difficult. The lower third of performers demonstrated almost no ability with the phoneme blending skill. Performance on items Q46 to Q55 which assess the ability to break words apart, rather than blend phonemes, indicated phoneme segmentation continued to be a difficult skill for most participants. This was not a surprising finding as Goswami and Bryant (1990) suggest that children progress through three levels of phonological awareness during their preschool and

early school years with phoneme level awareness being the final level of phonological awareness to develop.

General Item Analysis Conclusions

Phoneme level PA abilities are a crucial predictor of later reading skills in many children (Anthony et al., 2002). However, the challenges presented when assessing the skills of students with ASD, especially preschoolers, may mitigate some of the usefulness of this information (Koegel, Koegel, & Smith, 1997). Time, especially the perception of the use of time, is a valuable resource to educators (Collinson, 2001). Therefore, due to the assessed level of PA skills of preschoolers with ASD in this study, and to the behavioral challenges that may be associated with the high difficulty level items with this population, it may be warranted to at the very least shorten the phoneme level subtest to reduce testing time. While information tracking progress is important, it should be done as efficiently and as quickly as possible so that more time is left for instructing students.

Convergent and Discriminant Validity

The correlation analysis between the SBPAA and other measures revealed that this PA progress monitoring tool is highly correlated with the TOPEL. The correlations between post-test measures of PA (measured by the TOPEL) and the later administrations of both forms A and B of the SBPAA (r=0.86 and 0.89, respectively) were stronger than the correlations between the pre-test PA measures (TOPEL) and earlier administrations of forms A and B. Post-test correlation measures met Crocker and Algina's (1986) suggested threshold of r = 0.80 to 0.90. Pre-test correlation measures did not exceed this threshold, but they came close (r = 0.69 for form A and 0.77 for form B. In light of the fact that intervention specifically targeted these skills, it is not surprising that both measures would have higher levels of convergent validity after

intervention than before. Some of this difference may be due to the fact that subjects had received 7 lessons targeting PA before administration of form A and 29 lessons before administration of form B at post-test. While not ideal, the constraints of the dataset required that convergent validity was measured rather than concurrent validity.

The SBPAA also had relatively high levels of correlation with the *Test of Print Knowledge*. This is not surprising, as there is a bidirectional relationship between letter knowledge and phonological awareness (Lerner & Lonigan, 2016). Furthermore, Lonigan and colleagues (2009) posited that this relationship may exist because phonological memory and phonological awareness may be part of the same construct in preschool children. This suggests that as capacities in one area are expanded, they likewise grow in the other.

In regard to discriminant validity, no oral language measure produced highly divergent scores from administrations of the SBPAA. This is also not surprising, as the National Early Literacy Panel (2008) found that composite measures that addressed both expressive and receptive language correlated at a level of .70 with reading comprehension and .58 with decoding (Shanahan & Lonigan, 2013). Furthermore, Kurdek and Sinclair (2000) found that measures of verbal readiness skills in Kindergarten were highly predictive of multiple academic outcomes in later grades. Finally, non-reading measures such as cognitive abilities were not available for analysis to control for variance within the dataset. Various cognitive-linguistic measures, such as rapid naming, have been found to substantially improve prediction of future reading and spelling skills for children with reading disabilities (Scarborough, 1998). In sum, many of the skills measured in this study are interrelated, thus high levels of correlation are to be expected.

Ability to Detect Change.

A repeated-measures design was used to compare scores from multiple assessments from the same individuals. Paired t-tests were used to compare total test scores between administrations of both forms of the SBPAA. The paired t-test of form A, administrations 1 and 4, yielded a statistically significant result ($p<.001$, $d = 1.31$) with an effect size that was much larger than typical (Leech, Barret, Morgan, 2008). The paired t-test of form B, administrations 2 and 3, also resulted in a statistically significant result ($p<.001$, $d = 0.99$) with a large effect size (Cohen, 1988). These results indicate that both forms of the SBPAA are adequately sensitive to detect change and growth of PA skills. Furthermore, it is encouraging to see that form B was able to detect change between lessons 30 and 46, a total of 16 lessons. Whereas analysis of form A compared change between lessons 8 and 66, a span of 58 lessons. This indicates the SBPAA is sensitive enough for use as a progress monitoring tool at least every 15 lessons, thus enabling a teacher to check if a student is responding to the interventions as they are being provided. If a student is not showing growth it might indicate the need to adjust teaching practices. ..

Limitations and Future Directions

Equivalence

Due to the nature of the dataset, statistical evidence for the equivalence of form A and form B is not available at this time. Future studies should consider concurrent administration of the two measures with a similar population in order to establish levels of equivalence through concurrent validity. Ideally, the sample would be large enough to allow for half of the study population to be administered form A first, and the second half of the population to be assessed using form B first (Allen & Yen 2002). Conducting a study in this way would allow the order of administration to be flipped and thus account for any order effects that might exist.

Possible Reliability Threats

The analyses regarding form A, administration 1, are preliminary and may need to be interpreted with caution. At this point in the study, seven lessons of intervention would have been provided. Most preschoolers, especially with developmental disabilities, would not be expected to know some of these skills. Because of the ability and skill levels of students in the sample, floor effects may have had an impact and inflated findings. However, floor effects are generally not considered until difficulty levels reach a level of 0.15. Therefore, in terms of psychometric properties, the majority of the assessment data would be considered free of inflation effects. However, caution in over-interpreting these findings is advised.

Validity Threats

Due to the nature of the dataset, this analysis examined the convergent validity between the SBPAA and the TOPEL. Ideally, concurrent validity would be measured using data from two measures that were administered at the same time rather than weeks apart. An additional confound to an ideal analysis was that intervention was provided between the two administrations.

Implications for Practice

Based on the findings of this preliminary study, the SBPAA appears to be a reliable and valid tool for the measurement and progress monitoring of the growth of PA skills in preschoolers with ASD. In regards to future use, this study is suggestive that it may be warranted to shorten the assessment and reduce the number of items intended to measure phoneme level skills. In this sample, while most of the items on the subtest met or approached necessary levels of reliability, difficulty, and discriminability, the items in the phoneme level section did not perform as consistently as items within other subtests. Furthermore, this skill is

often one of the last to develop (Goswami & Bryant, 1990) and thus may increase testing frustration with a population of children that are difficult to assess (Koegel, et al., 1997).

The Standards (APA, et al., 2014) call for evidence that a measure is reliable and valid for use with its intended population, especially if the test taker is not representative of a typical normative sample. This study adds evidence that the SBPAA is a suitable tool for use with preschoolers with ASD. The results collected while this sample of preschoolers with ASD underwent PA interventions (Hudson, et al., 2017) suggest that the SBPAA was a reliable measure, was able to discriminate between high and low performers, had high levels of convergent validity with the TOPEL, and was sufficiently sensitive to detect change as PA skills grow.

References

Adams, M. J. (1990). *Beginning to read : Thinking and learning about print*. Cambridge, MA: MIT Press.

Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Long Grove, IL: Waveland Press, Inc.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5.* (5th ed.). Washington, D.C.: American Psychiatric Association.

American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.), & American Educational Research Association. (2014). *Standards for educational and psychological testing*. (A. P. Association, N. C. on M. in Education, & J. C. on S. for E. and P. T. (U.S.), Eds.). Washington, DC: American Educational Research Association.

Anthony, J. L., Lonigan, C. J., Burgess, S. R., Driscoll, K., Phillips, B. M., & Cantor, B. G. (2002). Structure of preschool phonological sensitivity: Overlapping sensitivity to rhyme, words, syllables, and phonemes. *Journal of Experimental Child Psychology, 82*(1), 65-92.

Baron-Cohen, S. (1995). *Mindblindness : An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Bernier, R., & Gerdts, J. (2010). *Autism spectrum disorders : A reference handbook.* (Contemporary world issues). Santa Barbara, CA: ABC-CLIO.

Billstedt, E., Gillberg, C., & Gillberg, C. (2005). Autism after adolescence: Population-based 13- to 22-year follow-up study of 120 individuals with autism diagnosed in childhood. *Journal of Autism and Developmental Disorders*, *35*(3), 351–360.

Blachman, B. A. (2000). *Road to the code : A phonological awareness program for young*

children. Baltimore, MD: Paul H. Brookes.

Bonett, D., & Wright, T. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior, 36*, 3–15.

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., … Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*(4), 422–432.

Brassard, M. R., & Boehm, A. E. (2007). *Preschool assessment : Principles and practices*. New York, NY: Guilford Press.

Brown, H. M., Oram-Cardy, J., & Johnson, A. (2013). A meta-analysis of the reading comprehension skills of individuals on the autism spectrum. *Journal of Autism and Developmental Disorders*, *43*(4), 932–955.

Bryant, P., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology*, *26*(3), 429–438.

Byrne, B., Freebody, P., & Gates, A. (1992). Longitudinal data on the relations of word-reading strategies to comprehension, reading time, and phonemic awareness. *Reading Research Quarterly*, *27*(2), 140–151.

Catts, H., Hogan, T., & Fey, M. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities*, *36*(2), 151–164.

Centers for Disease Control and Prevention. (2007). Prevalence of autism spectrum disorders--Autism and Developmental Disabilities Monitoring Network, six sites, United States, 2000. *Morbidity and Mortality Weekly Report: Surveillance Summaries*, *56*(1), 12–28.

Centers for Disease Control and Prevention. (2009). Prevalence of autism spectrum disorders - Autism and Developmental Disabilities Monitoring Network, United States, 2006. *Morbidity and Mortality Weekly Report: Surveillance Summaries*, *58*(10), 1–20.

Centers for Disease Control and Prevention. (2014). Prevalence of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, *63*(2), 1–21.

Chard, D. J. (1998). Word recognition: Research bases. In D. C. Simmons & E. J. Kamenui (Eds.), *What reading research tells us about children with diverse learning needs: Bases and basics* (pp. 141–168). Mahwah, NJ: Erlbaum.

Chard, D. J., & Dickson, S. V. (1999). Phonological awareness: Instructional and assessment guidelines. *Intervention in School and Clinic*, *34*(5), 261–270.

Christensen, D. L., Baio, J., Braun, K. V. N., Bilder, D., Charles, J., Constantino, J. N., … Yeargin-Allsopp, M. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *Morbidity and Mortality Weekly Report. MMWR.* Atlanta, GA

Clarke, P., Snowling, M., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading comprehension difficulties: A randomised control trial. *Psychological Science, 21,* 1106–1116.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Collinson, V. (2001). "I don't have enough time": Teachers' interpretations of time as a key to learning and school change. *Journal of Educational Administration, 39*(3), 266-281.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*(2), 394–409.

Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M., & McLennan, J. (1989). Autism diagnostic interview: A standardized investigator-based instrument. *Journal of Autism and Developmental Disorders, 19*(3), 363–387.

Coyne, M. D., Kame'enui, E. J., & Simmons, D. C. (2004). Improving beginning reading instruction and intervention for students with LD: reconciling "all" with "each". *Journal of Learning Disabilities, 37*(3), 231–239.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt Brace Javanovich.

Cronbach, L. J. (1951). Coefficient alpha and the interval structure of tests. *Psychometrika, 16*, 297–334.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test Validity* (pp. 3–17). Hillsdale, NJ: Erlbaum.

Crowder, R. G. (1982). *The psychology of reading*. New York, NY: Cambridge University Press.

Davidson, M. M., & Ellis Weismer, S. (2014). Characterization and prediction of early reading abilities in children on the autism spectrum. *Journal of Autism and Developmental Disorders, 44*(4), 828–845.

DeVellis, R. (2006). Classical test theory. *Medical Care, 44*(11), 50–59.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446.

Dunn L. M., Dunn D. M. (2007). Peabody picture vocabulary test, fourth edition (PPVT-4). Minneapolis, MN: NCS Pearson

Donnellan, A. M. (1985). *Classic readings in autism*. (A. M. Donnellan, Ed.). New York, NY: Teachers College Press.

Dynia, J. M., Lawton, K., Logan, J., & Justice, L. M. (2014). Comparing emergent-literacy Skills and home-literacy environment of children with autism and their peers. *Topics in Early Childhood Special Education, 34*(3), 142-153.

Eigsti, I. M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, *5*(2), 681–691. http://doi.org/10.1016/j.rasd.2010.09.001

Estes, A., Rivera, V., Bryan, M., Cali, P., & Dawson, G. (2011). Discrepancies between academic achievement and intellectual ability in higher-functioning school-aged children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*(8), 1044–1052. http://doi.org/10.1007/s10803-010-1127-3

ESSA (2015). Every Student Succeeds Act of 2015, Pub. L. No. 114-95 § 114 Stat. 1177 (2015-2016).

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education*, *41*(2), 121–139.

Fombonne, E. (2003). The prevalence of autism. *JAMA, 289*(1), 87–89.

Foorman, B. R., & Torgesen, J. K. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice*, *16*(4), 203–212.

Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing*, *18*(2), 129–155.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *Journal of Learning Disabilities*, *44*(4), 339–347.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78*(3), 263–279.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice*, *18*(3), 157–171.

Fuchs, L. (2004). The, past, present, and future of curriculum-based measurement research. *School Psychology Review*, *33*, 188–192.

Fuchs, L. S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children*, *39*(5), 14–20.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities*, *45*(3), 195–203.

Gabig, C. S. (2010). Phonological awareness and word recognition in reading by children with autism. *Communication Disorders Quarterly*, *31*(2), 67–85.

Garretson, H. B., Fein, D., & Waterhouse, L. (1990). Sustained attention in children with autism. *Journal of autism and developmental disorders*, *20*(1), 101-114

George, D., & Mallery, P. (2003). *SPSS for windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Geschwind, D. H., & Levitt, P. (2007). Autism spectrum disorders: developmental disconnection syndromes. *Current Opinion in Neurobiology*, *17*(1), 103–111.

Gillon, G. T. (2004). *Phonological awareness : From research to practice*. New York, NY: Guilford Press.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read (Essays in developmental psychology)*. Hove: Erlbaum.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10.

Griffin, P., Burns, M., & Snow, C. (1998). *Preventing reading difficulties in young children*. National Academy Press.

Griswold, D. E., Barnhill, G. P., Myles, B. S., Hagiwara, T., & Simpson, R. L. (2002). Asperger syndrome and academic achievement. *Focus on Autism and Other Developmental Disabilities*, *17*(2), 94–102.

Haager, D., & Windmueller, M. P. (2001). Early reading intervention for english language learners at-risk for learning disabilities: Student and teacher outcomes in an urban school. *Learning Disability Quarterly*, *24*(4), 235–250.

Hammill, D. (2004). What we know about correlates of reading. *Exceptional Children*, *70*(4), 453–468.

Harrington, J. W., & Allen, K. (2014). The clinician's guide to autism. *Pediatrics in Review / American Academy of Pediatrics*, *35*(2), 62–78.

Harvard Health Publications. (2010). Are rates of autism spectrum disorders increasing? *The Harvard Mental Health Letter*.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement & Evaluation in Counseling & Development (American Counseling Association)*, *34*(3), 177.

Hiebert, E. H. (2009). *Reading more, reading better*. (E. H. Hiebert, Ed.). New York, NY: Guilford Press.

Hoover, W., & Gough, P. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, *2*, 127–160.

Huang, F. L. (2014). Using a bifactor model to assess the factor structure of the phonological awareness literacy screening for grades 1 through 3. *Journal of Psychoeducational Assessment*, *32*(7), 638–650.

Hudson, R.F., Sanders, E.A., Greenway, R., Xie, S., Smith, M., Gasamis, C., Martini, J., Schwartz, I., & Hackett, J. (in press). Effects of emergent literacy interventions for preschoolers with ASD. *Exceptional Children.*

Huemer, S. V., & Mann, V. (2010). A comprehensive profile of decoding and comprehension in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *40*(4), 485–493.

Hulme, C., & Snowling, M. (2009). *Developmental disorders of language learning and cognition*. Oxford, UK: Wiley-Blackwell.

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004)

Johnson, B., & Christensen, L. (2012). *Educational research: quantitative, qualitative, and mixed approaches* (4th ed.). Thousand Oaks, CA: Sage Publications.

Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement, 38*(4), 319–342.

Kane, M. (2004). Certification testing as an illustration of argument-based validation. *Measurement, 2*(3), 135–170.

Kane, M. (2006). Validation. In R. L. Linn & American Council on Education (Eds.), *Educational Measurement* (4th ed., pp. 17–64). New York, NY: Macmillan Publishing.

Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73.

Kapp, S. K., Gillespie-Lynch, K., Sherman, L. E., & Hutman, T. (2012). Deficit, difference, or both? Autism and neurodiversity. *Developmental Psychology, 49*(1), 59–71.

Koegel, L.K., Koegel, R.L., & Smith, A. (1997). Variables related to differences in standardized test outcomes for children with autism. *Journal of Autism and Developmental Disorders, 27*(3), 233-43.

Koritsas, S., & Iacono, T. (2011). Secondary conditions in people with developmental disability. *American Journal on Intellectual and Developmental Disabilities, 116*(1), 36–47.

Kurdek, L., & Sinclair, R. (2000). Psychological, family, and peer predictors of academic outcomes in first- through fifth-grade children. *Journal of Educational Psychology, 92*(3), 449-457.

Lane, H. B., Pullen, P. C., Eisele, M. R., & Jordan, L. (2002). Preventing reading failure: Phonological awareness assessment and instruction. *Preventing School Failure: Alternative Education for Children and Youth, 46*(3), 101–110.

Lane, H., & Pullen, P. (2004). *Phonological awareness assessment and instruction: A sound beginning*. Boston, MA: Pearson.

Lanter, E., Watson, L. R., Erickson, K. A., & Freeman, D. (2012). Emergent literacy in children with autism: An exploration of developmental and contextual dynamic processes. *Language, Speech, and Hearing Services in Schools, 43*(3), 308-324.

Leech, N., Barrett, K., & Morgan, G. (2008). *SPSS for intermediate statistics: Uses and Interpretation* (3rd ed.). New York, NY: Routledge.

Lerner, & Lonigan. (2016). Bidirectional relations between phonological awareness and letter knowledge in preschool revisited: A growth curve analysis of the relation between two code-related skills. *Journal of Experimental Child Psychology, 144*, 166-183.

Lewkowicz, N. K. (1980). Phonemic awareness training: what to teach and how to teach it. *Journal of Educational Psychology, 72*(5), 686–700.

Liss, M., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., Waterhouse, L., & Rapin, I. (2001). Executive functioning in high-functioning children with autism. *Journal of Child Psychology and Psychiatry, 42*(2), 261-270.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children : Evidence from a latent-variable longitudinal study. *Developmental Psychology*, *36*(5), 596–613.

Lonigan, C. J., Wagner, R. K., & Torgesen, J. K. (2007). *Test of preschool early literacy*. Austin, TX: ProEd.

Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lotter, V. (1966). Epidemiology of autistic conditions in young children. *Social Psychiatry, 1*(3), 124–135.

Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York, N.Y.: Routledge.

Martin N. A., Brownell R. (2011). Expressive one-word picture vocabulary test, fourth edition. Novato, CA: Academic Therapy

Matson, J. L., & Kozlowski, A. M. (2011). The increasing prevalence of autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(1), 418–425.

Messick, S. (1989). Validity. In R. L. Linn & National Council on Education and Measurement (Eds.), *Educational Measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan Publishing.

Meyer, J. (2010). *Reliability (Series in understanding measurement)*. New York, NY: Oxford University Press.

Minshew, N., Goldstein, G., Taylor, H. G., & Siegel, D. (1994). Academic achievement in high functioning autistic individuals. *Journal of Clinical and Experimental Neuropsychology*, *16*(2), 261–270.

Montes, G., & Halterman, J. S. (2008). Child care problems and employment among families with preschool-aged children with autism in the United States. *Pediatrics*, *122*(1), e202.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, *40*(5), 665–681. http://doi.org/10.1037/0012-1649.40.5.665

Nation, K., Clarke, P., Wright, B., & Williams, C. (2006). Patterns of reading ability in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *36*(7), 911–919.

Nation, K., Cocksey, J., Taylor, J., & Bishop, D. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry*, *51*, 1031–1039.

Nation, K., & Norbury, C. (2005). Why reading comprehension fails: Insights from developmental disorders. *Topics in Language Disorders*, *25*(1), 21–32.

National Center for Health Statistics (U.S.), & Blumberg, S. J. (2013). *Changes in prevalence of parent-reported autism spectrum disorder in school-aged U.S. children: 2007 to 2011-2012*. (National Center for Health Statistics (U.S.), Ed.). Hyattsville, MD : U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the national early literacy panel.* (N. C. for F. L. (U.S.), Ed.) *Report of the National Early Literacy Panel*. Washington, D.C. : National Institute for Literacy.

National Institute of Child Health and Human Development (U.S.). (2005). *Autism overview what we know.* (National Institute of Child Health and Human Development (U.S.), Ed.). Rockville, MD: U.S. Dept. of Health and Human Services, National Institutes of Health, National Institute of Child Health and Human Development.

National Reading Panel, & National Institute of Child Health and Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington D.C.: U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Child Health and Human Development.

Neuman, S. B., & Dickinson, D. K. (2001). *Handbook of early literacy research*. New York: Guilford Press.

Newschaffer, C. J., Falb, M. D., & Gurney, J. G. (2005). National autism prevalence trends from United States special education data.. *Pediatrics*, *115*(3), 783.

No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002).

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*(1), 1–18. http://doi.org/10.1016/0022-2496(66)90002-2

Oakhill, J., & Cain, K. (2007). Introduction to comprehension development. In *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 3–40). New York, NY: Guilford Press.

Pallant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using spss for windows* (3rd ed.). Maidenhead: Open University Press.

Pennington, B. F., & Bishop, D. V. M. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology*, *60*, 283–306. http://doi.org/10.1146/annurev.psych.60.110707.163548

Phillips, B. M., & Piasta, S. B. (2013). Phonological awareness and alphabet knowledge: Key precursors and instructional targets to promote reading success. In T. Shanahan & C. J. Lonigan (Eds.), *Early Childhood Literacy: The National Early Literacy Panel and Beyond* (pp. 95–116). Baltimore, MD: Brookes.

Hill, A. P., Zuckerman, K., & Fombonne, E. (2015). Epidemiology of autism spectrum disorders. In *Translational Approaches to Autism Spectrum Disorder* (pp. 13-38). Springer International Publishing.

Pullen, P., Lane, H., & Hayes, L. (1999). *Jump start in reading: Assessments of early literacy development*. Gainesville, FL.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Ed). Thousand Oaks, CA: Sage Publications.

Rice, C. E., Rosanoff, M., Dawson, G., Durkin, M. S., Croen, L. A., Singer, A., & Yeargin-Allsopp, M. (2012). Evaluating changes in the prevalence of the autism spectrum disorders. *Public Health Reviews*.

Ricketts, J. (2011). Research review: Reading comprehension in developmental disorders of language and communication. *Journal of Child Psychology and Psychiatry*. Oxford, UK.

Rupley, W., & Willson, V. (1997). Relationship between reading comprehension and components of word recognition: Support for developmental shifts. *Journal of Research and Development in Education, 30*(4), 255–260.

Sandall, S. R., Schwartz, I. S., & Lacroix, B. (2004). Interventionists' perspectives about data collection in integrated early childhood classrooms. *Journal of Early Intervention*, *26*(3), 161–174.

Sansosti, F. J., Powell-Smith, K. A., & Cowan, R. J. (2010). *High-functioning autism/asperger syndrome in schools : Assessment and intervention*. (K. A. Powell-Smith & R. J. Cowan, Eds.). New York, NY: Guilford Press.

Scarborough, Hollis S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia, 68,* 115-36.

Shanahan, T., & Lonigan, Christopher J. (2013). *Early childhood literacy: The national early literacy panel and beyond*. Baltimore, MD.: Paul H. Brookes Pub.

Shaywitz, S., Escobar, M., Shaywitz, B., Fletcher, J., & Makugh, R. (1994). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*, *326*(3), 145–150.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *3*, 1–18.

Stanovich, K. Cunningham, A. & Cranner, B. (1984). Assessing phonological awareness in kindergarten children: Issue of task comparability. *Journal of Experimental Child Psychology*, *38*, 175–190.

Stecker, P. M., Lembke, E. S., & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision making. *Preventing School Failure: Alternative Education for Children and Youth*, *52*(2), 48–58.

Stoelb, M., Yarnal, R., Miles, J., Takahashi, T. N., Farmer, J. E., & McCathren, R. B. (2004). Predicting responsiveness to treatment of children with autism: A retrospective study of the importance of physical dysmorphology. *Focus on Autism and Other Developmental Disabilities*, *19*(2), 66–77.

Stoolmiller, M., & Bank, L. (1995). Autoregressive effects in structural equation models: We see some problems. In J. Gottman (Ed.), *The analysis of change* (pp. 261–276). Hillsdale, NJ: Erlbaum.

Storch, S. A, & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental Psychology*, *38*(6), 934–947.

Tager-Flusberg, H., & Caronna, E. (2007). Language disorders: Autism and other pervasive developmental disorders. *Pediatric Clinics of North America*, *54*(3), 469–481.

Traub, R. (1997). Classical test theory in historical perspective. *Educational Mesurement: Issues and Practice*, *16*(4), 8–14.

VanTilborg, A., Segers, E., van Balkom, H., & Verhoeven, L. (2014). Predictors of early literacy skills in children with intellectual disabilities: A clinical perspective. *Research in Developmental Disabilities*, *35*(7), 1674–1685. http://doi.org/10.1016/j.ridd.2014.03.025

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*(2), 192–212.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology*, *30*(1), 73–87.

Wagner, R., Torgesen, J., Rashotte, C., Hecht, S., Barker, T., Burgess, S., … Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: a 5-year longitudinal study. *Developmental Psychology*, *33*(3), 468–479.

Webber, J., & Scheuermann, B. (2008). *Educating students with autism : A quick start manual*. (B. Scheuermann, Ed.). Austin, TX: PRO-ED.

Weiner, I., & Greene, R. (2008). *Handbook of personality assessment*. New York, NY: Wiley.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *69*(3), 848–872.

Whitehurst, G., & Lonigan, C. (2001). Emergent literacy: Development from prereaders to readers. In *Handbook of early literacy research* (pp. 11–29).

Williams, K., Mellis, C., & Peat, J. K. (2005). Incidence and prevalence of autism. *Advances in Speech Language Pathology*, *7*(1), 31–40.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III tests of achievement. Itasca, IL: Riverside Publishing.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly*, *23*(2), 159–177.

Zirkel, P. a, & Thomas, L. B. (2010). State laws for RTI: An updated snapshot. *Council for Exceptional Children*, *42*(3), 56–63.

Appendix A

SBPAA Form A

Skill: Tapping Words

Directions: Say words in bold to student. **In this word game, I want you to tap one time for every word you hear in my sentence. So, if I say, "*An apple is red,*" you tap like this."** [Model saying the sentence tand tapping 4 times.] **Now you try.** Repeat with practice item and provide corrective feedback as needed. Do not provide corrective feedback on the test items.

Practice Item: "The little frog is jumping"

Test Items:

1. Sue is my friend. 4 taps _____
2. My mother is calling me. 5 taps _____
3. The dog is chasing the kittens. 6 taps _____
4. The boy is running. 4 taps _____
5. She is wearing a purple jacket. 6 taps _____

Word Tapping Score _____

Skill: Deleting Words

Practice Item: Say COWBOY. Now say COWBOY without saying BOY.

Test Items:

1. Say SAILBOAT. Now say SAILBOAT without saying SAIL. BOAT _____
2. Say APPLESAUCE. Now say APPLESAUCE without saying SAUCE. APPLE _____
3. Say SIDEWALK. Now say SIDEWALK without saying SIDE. wALK _____
4. Say SUNSHINE. Now say SUNSHINE without saying SUN. SHINE _____
5. Say CUPCAKE. Now say CUPCAKE without saying CAKE. CUP _____

Word Deletion Score _____

Total Word Level Score _____

Skill: Blending Syllables

Directions: For each item, say each word with a clearly discernible pause between syllables. **I'm going to say a word one part at a time. I want you to listen carefully, then put the parts together to make a whole word. Ready? Lets try one.**

Practice Item: What word do these sounds make? 'can-dy' [candy]

Test Items:

1. tea-cher teacher _____
2. hopp-ing hopping ____
3. va-ca-tion vacation ____
4. mo-tor-cy-cle motorcycle ____
5. re-fri-ger-a-tor refrigerator ____

Blending Syllables Score ____

Skill: Tapping Syllables

Directions: Follow the procedures for tapping words in the previous section. Begin by modeling how to tap the syllable in the child's name. **We can tap the parts of your name.** [Model using child's name.] **Now you tap the parts in your name. Let's try another one.**

Practice item: COMPUTER [3 taps]

Test Items:

1. ALLIGATOR 4 taps ____
2. PAPER 2 taps ____
3. JUMP 1 tap ____
4. DINASAUR 3 taps ____
5. BASKET 2 taps ____

Tapping Syllables Score ____

Skill: Deleting Syllables

Practice Item: Say WONDER. Now say WONDER without saying DER.

Test items:

1. Say SANDWICH. Now say SANDWICH without saying SAND. (WICH) _____
2. Say PENCIL. Now say PENCIL without saying CIL. (CIL) _____
3. Say ELBOW. Now say ELBOW without saying EL. (EL) _____
4. Say ANSWERING. Now say ANSWERING without saying ING (ANSWER) _____
5. Say REMEMBER. Now say REMEMBER without saying RE. (MEMBER) _____

Deleting Syllables Score _____

Total Syllable Level Score _____

Skill: Matching Rhymes

Practice Items: Do these two words rhyme: sack/black? Do these two words rhyme: beat/bean?

Test items;

1. Do these two words rhyme? hill/still Yes _____
2. Do these two words rhyme? drip/drop No _____
3. Do these two words rhyme? jump/bump Yes _____
4. Do these two words rhyme? dock/rock Yes _____
5. Do these two words rhyme? day/cake No _____

Skill: Blending Onsets and Rimes

Practice Item: What word do these sounds make? n-ote

Test items:

1.	f-an	fan	_____
2.	s-ip	sip	_____
3.	c-at	cat	_____
4.	st-op	stop	_____
5.	d-ance	dance	_____

Blending Onsets and Rimes Score _____

Skill: Generating Rhymes

Practice Items: Tell me a word that rhymes with *sat.* (Accept any rhyming word or pseudoword.)

Test items:

1. Tell me a word that rhymes with *mop.* ____
2. Tell me a word that rhymes with *run.* ____
3. Tell me a word that rhymes with *name.* ____
4. Tell me a word that rhymes with *tip.* _____
5. Tell me a word that rhymes with *sell.* _____

Skill: Blending Phonems

Directions: Stop administering when child misses 5 items in a row. Segment sounds as indicated by a (-) in each item. Be sure to pause briefly but discernibly between segmented sounds. If the child asks you to repeat the sounds, you may do so.

If the child says sounds separately (e.g., m-e, rather than me) prompt by saying, **Try to say the sounds altogether as a real word.** This prompt may be used as often as needed on practice items only.

Practice Items: **I'm going to say a word in small parts. I will say one part of the word at a time. I want you to listen carefully, then put the parts together to make a whole word. Ready? Lets try one. What word do these sounds make? 'l-i-p'**

If correct: **That's right. Let's try the next one.**
If incorrect: **That's not quite right. When you put 'l-i-p' together, it makes 'lip.' Let's try the next one.**

Continue to give corrective feedback as above. **What word do these sounds make? i-t? What word do these sounds make: m-a-t?**

Let's try some more words. Each time, I will say a word one part at a time. Listen carefully and put the parts together to make a whole word.

Test Items: (provide no more feedback)

1.	n-o	no _____
2.	i-f	if _____
3.	sh-e	she _____
4.	s-u-n	sun _____
5.	c-a-p	cap _____
6.	b-o-th	both ____
7.	f-a-s-t	fast _____
8.	j-u-m-p	jump _____
9.	s-t-ar-t	start _____
10.	s-p-l-a-sh	splash _____

Blending Phonemes Score _____

Skill: Segmenting Phonemes

Directions: Stop administering when child misses 5 items ina row. **I am going to say a word. I want you to break the word apart and say it sound by sound, like I did in the last word game we played. For example, if I say DAY, you would say it sound by sound, D-AY. Now you say that word DAY sound by sound.**

If correct: **That's right. You have the idea. Let's try the next one.**
If incorrect: **That's not quite right. To say DAY sound by sound, say D-AY. Let's try the next one.**

Practice Items: (continue corrective feedback)

Remember, I will say a word, then you say it sound by sound.

a.	no	n-o
b.	it	i-t
c.	pie	p-ie

Lets try a few more. I will say a word, then you say it sound by sound.

Test items: (no feedback)

1.	go	g-o _____
2.	up	u-p _____
3.	me	m-e _____

Directions: **Now we are going to do some words that have more sounds in them. Let's try a few for practice.** (Give corrective feedback as on previous practice times.)

Practice Items:

a.	mat	m-a-t
b.	good	g-oo-d
c.	winner	w-i-nn-er

Test items: (no feedback)

4.	back	b-a-k _____
5.	mop	m-o-p _____
6.	pig	p-i-g _____
7.	stop	s-t-o-p _____
8.	smoke	s-m-o-ke _____
9.	toast	t-oa-s-t _____
10.	mother	m-o-th-er ____

Segmenting Phonemes Score _____
Total Phoneme Level Score _____

Appendix B

SBPAA Form B

Skill: Tapping Words

Test Items:

1. Mike is my dog. 4 taps _____
2. My father is hugging me. 5 taps _____
3. The cat is eating the cookies. 6 taps _____
4. The girl is swimming. 4 taps _____
5. He is wearing a yellow blazer. 6 taps _____

Word Tapping Score _____

Skill: Tapping Words

Practice Item: Say COWBOY. Now say COWBOY without saying BOY.

Test Items:

6. Say AIRPLANE. Now say AIRPLANE without saying AIR. PLANE _____
7. Say MEATBALL. Now say MEATBALL without saying BALL. MEAT _____
8. Say MOONLIGHT. Now say MOONLIGHT without saying MOON. LIGHT _____
9. Say FIREWORKS. Now say FIREWORKS without saying FIRE. WORKS _____
10. Say POPCORN. Now say POPCORN without saying CORN. POP _____

Word Deletion Score _____

Skill: Blending Syllables

Directions: For each item, say each word with a clearly discernible pause between syllables.

I'm going to say a word one part at a time. I want you to listen carefully, then put the parts together to make a whole word. Ready? Let's try one.

Practice: What word do these sounds make? Can-dy

Test Items:

1. Blea-cher Bleacher _____
2. Jump-ing Jumping _____
3. Ex-er-cise Exercise _____
4. El-ev-a-tor Elevator _____
5. Hip-po-pot-a-mus Hippopotamus ____

Skill: Tapping Syllables

Directions: We can tap the parts of your name. Now you tap the parts in your name. Let's try another one.

Practice COMPUTER (3 taps)

Test Items:

1. Helicopter 4 taps _____
2. Monkey 2 taps _____
3. Run 1 tap _____
4. Elephant 3 taps _____
5. Monster 2 taps _____

Tapping Syllables Score _____

Skill: Tapping Syllables

Practice: Say Wonder. Now say Wonder without saying Der.

Test Items:

1. Airport without port Air ___
2. Picnic without nic Pic ___
3. Candle without can Dle ___
4. Supporting without ing Support ___
5. December without de Cember ___

Skill: Matching Rhymes

Practice Items: Do these two words rhyme? Sack/black

Test Items:

1. Mill/bill Yes ____
2. Brim/brom No _____
3. Hump/lump Yes ____
4. Sock/lock Yes ____
5. Bay/make No _____

Skill: Generating Rhymes

Practice Item: Tell me a word that rhymes with sat.

Test Items:

1. Tell me a word that rhymes with hip. ____
2. Tell me a word that rhymes with tan. ____
3. Tell me a word that rhymes with line. ____
4. Tell me a word that rhymes with bake. ____
5. Tell me a word that rhymes with bed. ____

Skill: Blending Phonemes

Directions: Stop when child misses 5 items in a row. Segment sounds as indicated by a (-) in each item. Be sure to pause briefly but discernibly between segmented sounds. If the child asks you to repeat, you may do so.

If the child says the sounds separately, prompt by saying **Try to say the sounds altogether as a real word.**

Practice Items: **I'm going to say a word in small parts. I will say one part at a time. You get to put the sounds together to make a whole word. Ready? What word do these sounds make? l - i – p**

If correct: **That's right. Let's try the next one.**

If incorrect: **That's not quite right. When you put l-i-p together, it makes lip.**

Do it again (with correct feedback) with m-a-t.

Test items:

1.	s-o	so ____
2.	i-n	in _____
3.	s-ee	see ____
4.	f-u-n	fun ____
5.	c-a-t	cat ____
6.	b-a-th	bath ___
7.	l-a-st	last ___
8.	d-a-mp	damp ____
9.	s-t-ar-k	stark _____
10.	s-p-l-o-sh	splosh _____

Skill: Segmenting Phonemes

Directions: Stop administering when the child misses 5 items in a row. **I am going to say a word. I want you to break the word apart and say it sound by sound, like I did in the last word game we played. For example, if I say DAY, you would say it sound by sound, D-AY. Now you say DAY sound by sound.**

If correct: **That's right. Let's do the next one.**

If incorrect: **That's not quite right. To say DAY sound by sound, say D-AY. Let's do the next one.**

Practice items (provide feedback):

a.	no	n-o
b.	it	i-t
c.	pie	p-ie

Test Items (no feedback):

1.	so	s-o _____
2.	am	a-m _____
3.	see	s-ee _____

Directions: Now we are going to do some words that have more sounds in them. Let's try a few more for practice. Give corrective feedback on the practice items.

Practice Items (provide feedback):

a.	mat	m-a-t
b.	good	g-oo-d
c.	winner	w-i-nn-er

Test items (no feedback):

4.	sick	s-i-ck _____
5.	hot	h-o-t _____
6.	bug	b-u-g _____
7.	step	s-t-e-p _____
8.	smite	s-m-i-te ____
9.	host	h-o-s-t _____
10.	father	f-a-th-er ____

Appendix C

Order of Administration of Forms A and B

	Administration			
Student	1	2	3	4
1	**A**	*B*	*B*	**A**
2	**A**	*B*	*B*	**A**
3	**A**	*B*	*B*	**A**
4	**A**			
5	**A**	*B*	*B*	**A**
6	**A**	*B*	*B*	**A**
7	**A**	*B*	*B*	**A**
8	**A**	*B*	*B*	**A**
9	**A**	*B*	*B*	**A**
10	**A**	*B*	*B*	**A**
11	**A**	*B*	*B*	**A**
12	**A**	*B*	*B*	**A**
13	**A**	*B*	*B*	**A**
14	**A**	*B*	*B*	**A**
15	**A**	*B*	*B*	**A**
16	**A**	*B*	*B*	
17	**A**	*B*	*B*	
18	**A**	*B*	*B*	
19	**A**	*B*	*B*	**A**
20	**A**	*B*	*B*	**A**
21	**A**	*B*	*B*	
22		*B*	*B*	
23			*B*	
24			*B*	
25			*B*	
26	**A**	*B*		
27	**A**	*B*		
28	**A**			
29	**A**		**A**	
30	**A**		**A**	
31	**A**	*B*		

32	**A**	*B*	**A**	
33	**A**	*B*	**A**	*B*
34	**A**	*B*	**A**	*B*
35	**A**	*B*	**A**	*B*
36	**A**	*B*		
37	**A**	*B*		
38	**A**	*B*	**A**	*B*
39	**A**	*B*	**A**	
40	**A**	*B*		
41	**A**		**A**	
42	**A**		**A**	
43	**A**		**A**	
44	**A**	*B*		