Development of an Integrative-Comprehension Imagery Scale for Children With and Without Autism

Marcy Willard

University of Denver

Follow this and additional works at: https://digitalcommons.du.edu/etd

Recommended Citation
https://digitalcommons.du.edu/etd/706

This Dissertation is brought to you for free and open access by the Graduate Studies at Digital Commons @ DU. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu.
DEVELOPMENT OF AN INTEGRATIVE-COMPREHENSION IMAGERY SCALE FOR CHILDREN WITH AND WITHOUT AUTISM

A Dissertation

Presented to

the Faculty of the Morgridge College of Education

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Marcy Willard

June 2013

Advisor: Gloria Miller, PhD
Abstract

Autism is a debilitating disorder (Yurov et al., 2007) that is diagnosed in 1 in 88 children in America (CDC, 2012). The autism population overwhelmingly performs weakest in reading comprehension as compared to other academic areas (Chiang & Lin, 2007; Minshew, 1994). This identified weaknesses is concerning because comprehension is understood in the literature as the most critical curricular area (Chiang & Lin, 2007). One potential reason for these comprehension problems could be impaired imagery.

Neuropsychology research has found that children with autism cognitively process imagery differently than typical children, due to their unique brain structures (Just, Cherkassky, Keller, & Minshew, 2004). This study sought to uncover whether or not children with autism show weaker abilities on imagery-related comprehension tasks than typical children. However, no instrument exists that is validated for this purpose.

Archived diagnostic data were examined to develop a scale to assess imagery-related comprehension skills. Data were analyzed from 71 children (N=71), with and without autism, aged 5-13. A four phase approach from Benson & Clark (1982) was used to develop the instrument. A multiple regression analytic method was used to see whether or not diagnosis was a significant predictor of scores on the imagery scale. Results were that children with autism scored significantly lower than typically developing children, controlling for the effects of IQ, age, and gender. Implications for practice are discussed.
Acknowledgements

Dr. Gloria Miller deserves endless accolades for her support and wisdom throughout this journey. Dr. Miller is a guiding light in the darkness and she has never once led me astray. Dr. Susan Hepburn provided critical oversight as an expert in autism and in the research study where the data were collected. The population of children and families who suffer from this disorder are well served by Dr. Hepburn’s wisdom and tireless efforts. I owe a debt of gratitude to all of the families at JFK Partners who gave their time, and to the children who worked in earnest on these assessments, in order to provide much needed answers to researchers. Dr. Daniel Brisson joined statistical expertise with a warm encouragement that has meant a great deal to me and to this research. Dr. Karin Dittrick-Nathan brought extensive assessment knowledge as well as moral support to this project. Special thanks go to Jennifer Albanes, my dissertation partner; I don’t know how anyone writes a dissertation without her. Jen, those times up at that special little cabin on American Way are memories that will last a lifetime.

Rick, my husband, has found the courage to jump into my dreams, even though I could not tell him where the path would lead. His personal strength and tenacity are a force with which it is a privilege to be a partner. The laughter of my awesome children, Brad and Brian, echoed in the hallway outside my office as these words made their way onto the page. Yes, kids, mom really is ‘almost done.’ My own mom has been a constant source of support to the boys and to me. As for my dad, I do not believe the dream would have been there without him. Special thanks to all my family who has supported me: Becky, Emily, Tracy, Ray, Hillary & Christian, and all the Willards…thank you!
# Table of Contents

Chapter One: Introduction and Study Purpose ................................................................. 1
   The Autism Epidemic .................................................................................................. 1
      ‘Autism Spectrum Disorders’ defined. ....................................................................... 3
   Integrative Comprehension defined ......................................................................... 6
   Imagery impacts integrative comprehension skills .................................................... 7
      Imagery defined. ........................................................................................................ 8
   Do children with autism have weaker imagery? ......................................................... 10
      Imagery Assessment ............................................................................................... 11
   Study Purpose ........................................................................................................... 12

Chapter Two: Literature Review ..................................................................................... 13
   Autism academic profile: Strong decoders. Poor Comprehenders. ......................... 14
   Imagery plays central role in comprehension ............................................................ 17
   Bridges between Imagery and Autism ...................................................................... 20
      Main Idea & the Weak Central Coherence Account ............................................... 21
      Integration & the Functional Under-Connectivity theory ...................................... 23
   Imagery Assessment ................................................................................................. 26
   A Model of Integrative-Comprehension Imagery ..................................................... 32
   Research questions ................................................................................................... 42

Chapter Three: Method .................................................................................................. 44
   Design ....................................................................................................................... 44
   Participants ............................................................................................................... 46
      Inclusion Criteria ................................................................................................. 47
      Final Sample Description ................................................................................... 50
   Instruments .............................................................................................................. 52
      Diagnostic instruments ....................................................................................... 52
      Intelligence tests ................................................................................................. 54
      Outcome Instruments ......................................................................................... 58
      Willard Imagery Observation Scales ...................................................................... 60
   Analysis Plan ............................................................................................................ 60
      Imagery Scale Development ............................................................................... 61
      Imagery Assessment Results ............................................................................... 61

Chapter Four: Imagery Scale Development ................................................................... 65
   Phase I: Planning ....................................................................................................... 67
   Phase II: Test Construction ...................................................................................... 73
   Phase III: Reliability ............................................................................................... 75
   Phase IV: Validity ..................................................................................................... 98

Chapter Five: Imagery Assessment Results .................................................................. 107
   Results of regression on WIOS Total ....................................................................... 112
      Results of regression on Main Idea Domain ......................................................... 117
      Results of Regression on Integration Domain ..................................................... 120
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sample Description</td>
<td>51</td>
</tr>
<tr>
<td>2. Imagery Domains Defined in the Literature</td>
<td>70</td>
</tr>
<tr>
<td>3. Pilot Sample Description</td>
<td>77</td>
</tr>
<tr>
<td>4. Scale Reliability</td>
<td>80</td>
</tr>
<tr>
<td>5. Item Revisions to Enhance Reliability</td>
<td>82</td>
</tr>
<tr>
<td>6. Domain Changes Based on Factor Structure</td>
<td>92</td>
</tr>
<tr>
<td>7. Factor Loadings for Version 3</td>
<td>94</td>
</tr>
<tr>
<td>8. Factor Analysis for Two Factor Model</td>
<td>97</td>
</tr>
<tr>
<td>9. Final Scale</td>
<td>98</td>
</tr>
<tr>
<td>10. Psychometric Properties</td>
<td>100</td>
</tr>
<tr>
<td>11. Assessment Results</td>
<td>109</td>
</tr>
<tr>
<td>12. Relationship between Diagnosis and WIOS Total Score</td>
<td>111</td>
</tr>
<tr>
<td>13. Regression of Diagnosis on WIOS Total Score</td>
<td>113</td>
</tr>
<tr>
<td>14. Regression of Diagnosis on the Main Idea Domain</td>
<td>116</td>
</tr>
<tr>
<td>15. Regression of Diagnosis on the Integration Domain</td>
<td>119</td>
</tr>
</tbody>
</table>
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Intersection of Imagery, Comprehension and Autism</td>
<td>12</td>
</tr>
<tr>
<td>2.</td>
<td>Bridges between Imagery and Autism</td>
<td>21</td>
</tr>
<tr>
<td>3.</td>
<td>A Model of Integrative-Comprehension Imagery</td>
<td>33</td>
</tr>
<tr>
<td>4.</td>
<td>Sample Selection Procedure</td>
<td>47</td>
</tr>
<tr>
<td>5.</td>
<td>Instrument Development Flowchart</td>
<td>66</td>
</tr>
<tr>
<td>6.</td>
<td>Pilot Sample</td>
<td>76</td>
</tr>
<tr>
<td>7.</td>
<td>Validation Sample</td>
<td>99</td>
</tr>
<tr>
<td>8.</td>
<td>Final Sample</td>
<td>108</td>
</tr>
</tbody>
</table>
Chapter One: Introduction and Study Purpose

This chapter is a discussion of the diagnostic profiles of autism as well as the current research on prevalence and potential causes for the disorder. Definitions of integrative-comprehension and imagery are provided. Finally, a list of autism specific comprehension deficits is considered with a focus on weakness areas that may have a basis in mental imagery.

The Autism Epidemic

Just two decades ago, autism was a rare disorder, afflicting fewer than 1 in 10,000 children in the United States (Kaye, del Mar Melero-Montes, & Jick, 2001). The general public often thought of Dustin Hoffman’s character in the movie, Rainman (1988), when they thought of autism. In those days, people found autism rare and fascinating, picturing Hoffman’s character who was able to count hundreds of matches by sight but did not know the price of a candy bar. This picture of autism is in stark contrast to our understanding today.

Today, autism afflicts 1 in 88 children (CDC, 2012). These recently reported prevalence rates represent a 23% increase from the 2009 report (when prevalence was 1:110), and a 78% increase from the 2007 report (prevalence was 1:150) (CDC, 2012). The incidence of autism is higher than AIDS, pediatric cancer, and diabetes combined (Autism Speaks, 2011). Today, many people know a child afflicted by this disease and awareness has become widespread.
Autism is a neurological disorder. The primary symptoms include: impairments in social interaction, delayed or disordered communication skills, and repetitive or stereotyped behaviors (American Psychiatric Association (APA), 2000). Boys with autism outnumber girls 5:1 (CDC, 2012). Children with autism show significant social deficits (APA, 2000). They tend to have difficulty understanding the feelings or perspectives of other people (Ozonoff & Miller, 1995). Children with autism struggle with communicating and making friends, which is generally the most pervasive aspect of the disorder (Ozonoff & Miller, 1995). Mental illness and lower IQ scores are not diagnostic features of the disorder, however (APA, 2000). Many children with autism, and more so today than 5-10 years ago, have intellectual abilities in the average range (CDC, 2012).

The cause for autism is unknown. However, autism is primarily a genetic disorder, based on a complex combination of chromosomal abnormalities and gene mutations (Christian, & Cook, 2004; Harmon, 2012; Sanders et al., 2012, Veenstra-VanderWeele,). Scientists have discovered over 100 genes associated with autism and it has an extremely high heritability rate of .80-.90 (Broad Institute of MIT and Harvard, 2012). However, the precise gene variants responsible for autism are unclear (Yurov et al., 2007). Research indicates that older fathers show increased odds of having a child born with autism; likely because sperm division occurs throughout life, and these continual DNA replications increase risk of gene mutation in older men (Harmon, 2012; Schubert, 2008). Scientists have recently found a significantly high number of children with autism who are born to mothers who are obese, or suffer from metabolic conditions during pregnancy (Harmon, 2012; Krakowiak, et al., 2012).
Even with the strong evidence for heritability, most cases of autism do not have a clearly identified genetic cause (Yurov et al., 2007). Genetic scientists who study autism report that, “but in the majority of cases (90–95%), the cause of autism remains unexplained” (Yurov, et al., 2007, p.521). For this reason, many are led to consider a combination of environment and genetics as potential causes. At one time, there was the belief that autism could be linked to the Measles, Mumps and Rubella vaccine (Wakefield, 2000). However, recent scientific research has uncovered flaws in these early studies and has refuted these claims (Kaye, del Mar Melero-Montes, & Jick, 2001). There is currently no ‘cure’ for autism; although, there are cases of children who have been alleviated of most of their symptoms. Researchers refer to these instances not as ‘cured’ but as cases of ‘optimal outcomes’ (Hepburn & Katz, 2010). Generally speaking, this disorder is debilitating, increasing rapidly in prevalence, and the ‘cause’ and ‘cure’ are largely unknown.

‘Autism Spectrum Disorders’ defined.

Due to the wide range of symptom profiles, the term Autism Spectrum Disorders (ASD) was created. The five disorders under the umbrella term of ASD include: Autistic Disorder, Asperger’s Disorder, Pervasive Development Disorder, Not Otherwise Specified (PDD-NOS), Childhood Disintegrative Disorder (CDD), and Rett Syndrome. All of these disorders are currently diagnosed using a specific set of symptom criteria defined in a book called the DSM-IV (published by the American Psychiatric Association). The latter two disorders are extremely rare (Corsello, 2005, Fombonne, 2002;). These disorders are quite distinct and unique from autism (Kurita, Osada, & Miyake, 2004; Mahlotra & Gupta, 2002). Rett Syndrome may no longer be included in
the upcoming revisions of the DSM-V because a genetic cause has been identified for this disorder (APA, 2012; Volkmar, Koenig, & State, 2005). Childhood Disintegrative Disorder will likely be collapsed into the autism category in the DSM-V (Volkmar, Koenig, & State, 2005). As such, CDD and Rett Syndrome will not be the focus of this paper.

Children with other ASD’s such as autistic disorder, PDD-NOS, Asperger’s Syndrome, share a set of common symptoms. Children with ASD’s tend to show a lack of social reciprocity (Constantino & Todd, 2000). Their speech is sometimes robotic or parroted (Katz & Hepburn, 2010). Children with ASD’s tend to have special interests (APA, 2002); these interests are either not appropriate developmentally, or they are obsessive in nature (APA, 2002). Children with an ASD often have restricted play behaviors, tending to use less imaginative, symbolic, or pretend play than typical children (Rutherford, Young, Hepburn & Rogers, 2007). A child with an ASD might be seen flipping a doll’s eyes, or staring at the wheels of a toy train, rather than engaging in interactive play with peers.

Throughout this paper, children with autism will be compared to children with ASD’s, and to Typically developing children. The distinction between autism and another ASD (such as Asperger’s and PDD-NOS) was carefully considered in this research. Children with autism and Asperger’s are diagnosed on a continuum. Those with more severe symptoms meet a cut-off for autism. Children with autism symptoms, but lesser severity, meet criteria for Asperger’s Disorder or for PDD-NOS. Diagnostic criteria also specify that children with Asperger’s Disorder do not have significant cognitive impairment (DSM-IV: 299.90, Asperger’s Disorder). Further, children with Asperger’s
Disorder *must not* have a marked language delay in early childhood (APA, 2000). Thus, these disorders are quite similar but there are distinct diagnostic differences. The children in this study were diagnosed by highly trained clinicians, using ‘gold standard’ diagnostic instruments. As such, the diagnoses used in this study can be seen as valid and reliable.

Thus, throughout this paper, “children with autism” refers to autistic disorder specifically, rather than Asperger’s or PDD-NOS. This delineation between autism and Autism Spectrum Disorders was utilized in similar studies (Conti-Ramsden, Simkin, & Botting, 2006). A precise explanation of the coding protocol used in this study; which follows a similar continuum used for the diagnoses, is provided in the Methods chapter.

Throughout this paper, children without disabilities are referred to as ‘neurotypicals;’ ‘typical,’ or ‘typically developing’ children. These terms are considered interchangeable ways of describing children without disabilities. It should be noted here that the researcher is aware that while children with autism are not ‘typical’ in many ways, they are individuals with both unique and special characteristics. Thus, this research focuses on understanding these characteristic traits as a basis for maximizing the educational experiences of children with autism.

*Comprehension is impaired in children with autism.*

Decades of research and diagnostic data have revealed surprisingly little about the unique academic profiles of children with autism (Chiang & Lin, 2007). There is an old adage in the field, “If you have met one child with autism…you have met one child with autism.” This phrase embodies autism’s complicated, far reaching, and variable symptomology. The education field is in a state of flux; overwhelmed with this growing population of children with a wide variety of educational needs.
Even though the academic strengths and weaknesses of this population vary significantly, one hallmark feature is overwhelmingly consistent. In children with autism, a strong reading ability (decoding) is often coupled with poor reading comprehension (O’Connor & Klein 2004; Snowling, 1986). This decoding-comprehension discrepancy means that children with autism can read fluently without gaining meaning from what they have read.

**Autism academic profile: good decoders, poor comprehenders**

Children with autism are often said to be ‘word calling,’ or reading without comprehension (Huemer & Mann, 2009). In spite of an adequate or high reading level, comprehension remains low. It is as if the words are seen as an algorithm to be decoded rather than a message to be understood. Even though this academic profile is well documented, research is inconclusive as to the basis for this problem (Chiang & Lin., 2007; Huemer & Mann, 2009; Mayes & Calhoun, 2008). The question remains, what is different, about children with autism that renders comprehension so difficult? Before this question can be answered, some definitions are necessary.

**Integrative Comprehension defined.**

Scholars have debated the definition of comprehension for decades, resulting in dozens of theoretical orientations (Rapp & van den Broek, 2005). Integrative-comprehension refers to the integration of two interrelated cognitive processes. One process that occurs during reading is the text-level processing which involves an understanding of the story structure, including elements like: setting, actions, characters, major events, cause-and-effect relationships, and the logical sequence of events (Strickland & Feeky, 1985). This form of comprehension refers to the process of reading.
Generally, this process is considered to involve the activation of background knowledge and is considered a lower-level of comprehension (Cain, Oakhill, & Bryant, 2004).

However, comprehension also involves the products of reading which include the construction of a mental representation of the story’s overall meaning (Cain, et al., 2004; Kintsch, 1994). Integrative comprehension is the interaction of process and products (Rapp, et al., 2005). Readers with good integrative-comprehension construct a coherent and holistic model of what they have read, in addition to the facts that are gained from text-level processes.

Kintsch (1988); a recognized thought leader in the comprehension field, proposed the construction-integration model. He describes the model thusly,

It combines a construction process in which a text base is constructed from the linguistic input as well as from the comprehender’s knowledge base, with an integration phase, in which the text base is integrated into a coherent whole. (p.165)

*Integrative-comprehension* is thus a complex web of related skills that interact in order for the reader to gain text level knowledge; as well as, to form a coherent mental model of the text.

**Imagery impacts integrative comprehension skills.**

Although researchers have reliable evidence that integrative-comprehension problems are prevalent in the autism population, the basis for these problems remains largely unknown (Chiang & Lin, 2007; O’Connor & Klein, 2004). Comprehension problems in autism can be complex and far reaching and this study could not possibly address all of them. The purpose of this study is to uncover only one important skill that
is believed to be necessary for integrative-comprehension: imagery. Thus, the focus of this research is fairly narrow in that the researcher is only interested in integrative-comprehension as facilitated by imagery.

A strong body of literature supports the efficacy of comprehension interventions which are designed to address imagery-related weaknesses (Bell & Binetto; Cain, 2009; Clark & Paivio, 1987; Gambrell & Bales, 1986; Oakhill & Patel, 1991). Although the connection between imagery and autism has scarcely been made, these imagery interventions have been tested and shown to be effective in studies on children with similar disabilities (Joffe, Cain, & Maric, 2007; Norbury & Bishop, 2002, & Tager-Flusenberg, 2006). Thus, the researcher seeks to understand whether or not mental imagery is an issue in the autism population with the hope that future researchers will test imagery related comprehension interventions in the autism population if indeed these deficits are found.

**Imagery defined.**

Before a discussion of integrative-comprehension imagery ensues, it is necessary to define imagery. Imagery is the ability to create a picture in one’s mind, a mental representation, or a ‘snap-shot’ of what is being heard or read. Rather than looking at a picture on the page or referencing an object in the flesh, mental imagery is the ability to create original pictures in the mind. Imagery is the skill of following along with a story, forming a mental picture of the information and then dynamically adjusting that picture to integrate any new story information. Imagery might be thought of as a mental movie or moving picture slide-show in the mind (Gambrell & Bales, 1986; Kelly, 2007).
Imagery is best explained by way of example. The reader herein can develop his or her own image of the text, while experiencing imagery first-hand. Consider the following example,

Do you know which is darker green, a frozen pea or a pine tree? Or what shape are Mickey Mouse’s ears? Or the hand in which the statue of liberty holds the torch? Most people report that they answer these questions by visualizing the named objects, which allows them to remember the information. (Kosslyn, Thompson and Ganis, 2006, p.3)

It may not be possible to answer questions about the statue of liberty, for example, without having an image in one’s mind of what the statue looks like. Mental imagery is so natural to human cognition; it can be difficult to conceptualize as an isolated process. Of course, an individual’s images depend to a large degree on one’s lived experience and cultural context. Thus, our unique images are memory deposits based on an individualized archived set of personal knowledge and experiences. The human brain forms images all the time, as a way of understanding, remembering, and comprehending information. That is, rather than reading to remember facts, readers also seek to develop their own mental images that bring context and meaning to that experience.

The concept of imagery as a cognitive process has been studied since the early 1900’s (Isaac & Marks, 1994). However, today Alan Paivio is recognized as the modern-day guru and pioneer in discovering the process by which images are integrated with language in the mind. ‘Dual Coding Theory’ postulates that all language input is processed in the brain by way of these mental representations in the form of images and words (Paivio, 1990). Paivio is also responsible for the ‘conceptual peg’ hypothesis
which is the idea that images serve as ‘pegs’ in the mind where ideas are anchored; in
order to, conceptualize them, ponder them, and remember them. Thus, *Imagery*, for the
purposes of this study, is a mental representation of meaning, in the form of a mental
picture, which is constructed in the reader’s or listener’s mind.

**Do children with autism have weaker imagery?**

Theory suggests that children with autism have weaknesses in getting the main
idea of a text, integrating story elements, and the understanding of sensory details.
Similarly, reading comprehension researchers have found that all of these impairments
are often related to weaknesses in visual imagery (Bell & Binetto, 2006; Cain, Oakhill, &
Bryant, 2004; Gambrell & Bales, 1986; Joffe, Cain, & Maric, 2007). This research begs
the question as to whether children with autism have a weaker ability to use imagery. If
this problem exists, it could be an underlying cause for impaired comprehension in the
autism population. Further, this research could uncover a new basis for some of the
hallmark autism symptoms. Thus, it is of primary importance now to determine how to
best measure imagery-related comprehension abilities in children with autism.

Imagery is a complex cognitive process that has been explored from many
different angles. The researcher on this study does not attempt to address all areas of
imagery. Rather, this study is an exploration of a small set of inter-related integrative-
comprehension skills that are shown in the literature to require mental imagery. Further,
this research targets the specific skills that are often impaired in the autism population
with a careful look into whether or not these identified weaknesses may have a basis in
impaired imagery.
Imagery Assessment

To date, few standardized measures of imagery exist. There are a variety of informal approaches, and some valid self-report questionnaires, which are described in the next section. However, the researcher on this study was not able to uncover any instruments that could be used to assess the imagery skills that are important for integrative-comprehension and are often impaired in autism.

Further, in this study, it was important to consider the potential for assessing imagery related comprehension skills using observation. Children who come to a diagnostic clinic go through a lengthy, and sometimes daunting, assessment process. The introduction of yet another assessment for imagery may impose an undue burden on clinicians and examinees. As such, in this study, the researcher put forth the bulk of the effort considering how to use data that are already collected during a routine diagnostic assessment in order to glean additional information about imagery in autism.

Assessing integrative-comprehension imagery in autism is complex. It is important to be clear that the researcher does not attempt to assess all areas of comprehension; all areas of imagery; or all deficit areas in autism. Rather, the study that follows will look at the intersection of these. As shown below in Figure 1, the goal of this work is to develop an assessment of the most important integrative comprehension skills that require imagery to perform well, with a careful eye to whether or not those skills may be specifically impaired in children with autism.
Study Purpose

The purpose of this study is to develop an integrative-comprehension imagery assessment and then to apply that assessment to children previously identified with autism as compared to a control group of neurotypical children (age 5-13). This study will provide insights into imagery assessment in specific domains that are suspected to be impaired in autism. This research will provide three potential contributions to the field: 1) a model for integrative-comprehension imagery will be developed; 2) an integrative-comprehension imagery assessment tool will be designed and validated; and 3) this assessment will be used to consider whether or not children with autism demonstrate significant differences on integrative-comprehension imagery tasks compared to typically developing children.
Chapter Two: Literature Review

This chapter is a review of the literature on the reading comprehension challenges in the autism population in consideration of whether these issues may have a basis in mental imagery. First, there is a review of the legal and ethical requirement for schools to provide evidence-based instruction for children with ASD’s. Next, there is a discussion of the research indicating that children with autism tend to have poor reading comprehension in spite of adequate reading fluency. Then, there is a review of the literature on existing mental imagery assessment techniques. Following, is the description of the theoretical model of integrative-comprehension imagery, used as the basis for scale development. The chapter concludes with the research questions that will be addressed in this study, all regarding the assessment of imagery in autism.

Educating children with autism

The new revisions to the Individuals with Disabilities in Education Improvement Act (IDEA, 2004; revised 2006, 2007, 2008) requires that educators provide access to the full range of evidence-based instructional opportunities, that would reasonably provide educational benefit to students with autism (NAC, 2009; Yell, 2006). Due to the rising numbers of children with the diagnosis, the schools have experienced an influx of children on the autism spectrum. School psychologists have a legal and ethical requirement to meet the needs of students in this population (Schwartz, Burns, & Davis, 2008).
Thus, educators are left with a daunting mission: what do children with autism need and how can the system provide it? Research has clearly shown that comprehension is a significant weakness area for children with ASD’s (Chiang & Lin, 2007; Mayes & Calhoun, 2008). Moreover, the education literature points to comprehension as “the most important academic skill” (Mastropieri & Scruggs, 1997, p.1) Chiang and Lin (2007), reviewed a wide variety of studies on comprehension interventions; concluding that, 

Reading instruction for children with autism has been underemphasized and has gained very little research attention. (p.260)

**Autism academic profile: Strong decoders. Poor Comprehenders.**

There is a strong body of literature to indicate this pervasive pattern in the academic profiles of children with autism: high reading level (decoding) coupled with a lack of understanding (comprehension) (Huemer & Mann, 2009). Chiang & Lin (2007) conducted a research review of 11 studies regarding the academic profiles of children with autism. They concluded that all of the studies identified comprehension as the primary area of weakness.

Significant comprehension problems occur in this population even when performance is strong in other areas (Chiang & Lin, 2007; Minshew, 1994). No explanation exists, to date, for these difficulties (Chiang & Lin, 2007). That is, researchers have not explained why children who are intelligent and academically competent in so many areas, struggle to understand what they hear and read.

This academic pattern in autism has been supported in profile research, which is a way of exploring diagnostic data in order to uncover trends in the unique cognitive
profiles of children with specific disabilities. Mayes and Calhoun (2008) studied the diagnostic profiles of 54 children with autism, aged 6-14. This study examined the performance of children with autism on achievement (using Wechsler Individual Achievement Test) and intelligence tests (using the Wechsler Intelligence Scales for Children). They found that in over 94% of the cases, scores on the comprehension tasks were lowest, as compared to all other academic and cognitive tasks (Mayes & Calhoun, 2008).

It is important to note that the comprehension tasks on the WISC intelligence test are not intended to assess reading comprehension; rather, the tests measure comprehension of social situations and every-day problem solving. There are measures of reading comprehension on the achievement test. Even when comprehension was measured in different ways on cognitive and achievement assessments, children with autism tended to score lowest on comprehension tests (Mayes, & Calhoun, 2008).

Similar results were uncovered by Minshew (1994), who studied the overall academic achievement of 54 students with autism spectrum disorders as compared to a matched control group of 41 neurotypical school-aged children (mean age=15.5). Diagnosis was determined using the Autism Diagnostic Interview (ADI) and the Autism Diagnostic Observation Scales (ADOS). Controlling for IQ, age, and SES, children with and without autism were tested using several measures of achievement to determine strengths and weaknesses. Minshew predicted that students with autism would score in the average or high range on tasks of reading mechanics or math computation. However; she hypothesized that children with autism would score relatively low on comprehension
tasks. The hypothesis held up in the study, indicating that comprehension was a demonstrated significant academic weakness for children with autism in the sample as compared to neurotypical controls.

Strengthening these findings and exploring reading comprehension in more detail, Huemer and Mann (2009) compared a large sample of 384 participants with autism spectrum disorders to a control group of 100 participants with dyslexia. The average age of the group with autism spectrum disorder was 10-11 years; for the dyslexic group, the average was 11 years. The study utilized a factor-analysis procedure for the standard scores (using Z-transformations) on nine different standardized reading assessments. As hypothesized, the general trend was for children with autism to show a strong or adequate ability on decoding tasks; whereas, comprehension was a demonstrated relative weakness. The children with dyslexia showed the opposite pattern: low decoding skills accompanied by stronger comprehension.

In a related study on reading skills utilizing standardized reading assessments, Nation, Clarke, Wright, and Williams (2006) studied reading ability in children with ASD. Participants included 41 children with autism, aged 6-15 (Mean=10 years), who were tested using four measures of reading ability. Three of the measures assessed reading accuracy; the other measure assessed comprehension. Nation et al. found a great deal of variability in reading skills. However, the pattern was clear: even children with autism who were accurate readers displayed a significant weakness in comprehension skills. Children with autism had reading fluency scores in the normal range but reading comprehension at least 1SD below the population mean (Nation et al., 2006). The
researchers claim a close resemblance between the autism group and the hyperlexic reading pattern [extremely poor comprehension relative to decoding abilities].

All of this research begs important questions. First, why do children with autism struggle with comprehension? Secondly, do some of the significant comprehension problems have a basis in mental imagery?

**Imagery plays central role in comprehension**

The importance of imagery for comprehension has long been demonstrated. Researchers have found that mental imagery improves learning, memory, problem solving, and overall comprehension (Cain, 2009, Clark & Paivio, 1987; Gambrell & Bales, 1986; Joffe, Cain & Maric, 2007; Oakhill & Patel, 1991; Pressley, 1977). Good comprehension is an essential skill for academic success (Joffe et al., 2007) and imagery is critical to comprehension (Gambrell & Bales, 1986).

Studies show that children who use imagery effectively are able to: remember story details, produce coherent narratives, and demonstrate improved comprehension overall (Bell, 1991a; Cain, 2009; Clark & Paivio, 1987). Imagery strategies can aid in oral language, following directions, and in understanding cause and effect relationships (Bell, 1991b). Research indicates that children with stronger imagery abilities can make inferences, monitor their own comprehension, detect incongruences, and accurately respond to questions about stories they have read (Cain, 2009). Mather and Goldstein (2008), explain, that weaknesses in imagery can manifest into, “extreme difficulty with reading comprehension” (319).
Gambrell and Bales (1986) conducted a study by identifying 124 subjects, in the 4\textsuperscript{th} and 5\textsuperscript{th} grades who were one or two years below grade level in reading. Students were randomly assigned to treatment groups. The intervention group received direct instruction in visual imagery; the control group received no such instruction. Three high imagery sentences and two low-imagery sentences were presented. An example of a high imagery sentence is, “The happy little girl was eating pink, fluffy cotton candy.” The students were read passages that varied as to the degree of inconsistencies in the text. The intervention group (imagery condition) showed significant improved performance on these comprehension tasks in comparison to the control group. The researchers concluded that teaching imagery can be an effective strategy to improve reading comprehension. Gambrell and Bales claim, “Research has provided converging evidence that mental imagery facilitates reading comprehension” (p.455).

Similar findings were discovered by Oakhill and Patel (1991), who studied a group of 192 students, aged 9-10 years. Children were matched on vocabulary and decoding ability. The groups were divided based on comprehension skills. Both of these groups were provided with a three-session mental imagery intervention. The intervention groups were then compared to a control group that received no imagery training. The researchers found that mental imagery training improved comprehension for the children with poor comprehension skills.

Joffé, Cain, and Maric (2007) studied the effects of explicit instruction in imagery on comprehension skills. They explained that imagery can be used to make “an integrated representation or a mental model of the text” (p. 649). Participants in the study included
nine children with Specific Language Impairments (SLI) and sixteen typical children. Ages were not given; however, the students were in primary school in the United Kingdom and were matched for age in the study. Children in the intervention group were provided with 30 minute sessions of imagery training over three weeks. During the sessions, the intervention group was reminded to, “make pictures” as they read. Comprehension was measured using several standardized cognitive and language measures. The results of the study were that children who were provided with the imagery training improved comprehension significantly in comparison to controls.

Bell and Lindamood have developed an innovative imagery training program that has been highly effective in improving improved the reading comprehension of children with a wide range of abilities and challenges (Bell 1991a, 1991b, 2006; Bell & Binetto, 2006). The promising results of this program corroborate evidence from a variety of studies, indicating that imagery interventions can significantly improve reading comprehension in children with a variety of deficits and challenges (Cain, 2009; Clark & Paivio, 1991; Gambrell & Bales, 1986; Joffe et al., 2007; Oakhill & Patel, 1991).

If indeed imagery is demonstrated to be a significant challenge for children with autism, imagery intervention could offer much promise in remediating comprehension problems which are both socially and educationally debilitating for children in this population. However, at this point, no research exists, to demonstrate that imagery is an issue in autism. As such, the first step in understanding imagery deficits in autism is an assessment issue. The question, then, is how to test imagery abilities of children with autism.
**Bridges between Imagery and Autism**

This study utilized a unique approach to assess imagery in autism. The precise skills that are known to be impaired in autism were compared with the known domains of imagery. Most of the skills that are shown in the literature to require imagery are also demonstrated deficit areas in children with autism. In order to assess this, two major theories of autism were used to form the domains of the instrument: the Weak Central Coherence Account and the Functional Under-Connectivity theory. This ‘bridge’ between imagery and autism is depicted in Figure 2 that follows.
Main Idea & the Weak Central Coherence Account

Children with autism struggle to understand the main idea of a narrative when they read (Diehl et al., 2006; Nuske & Baven, 2010). Nuske and Baven (2010) compared a group of 14 children with autism, aged 8-18, to a matched control group, uncovering a unique difference. They found that even when the children with autism knew the main idea of the story, they failed to use that knowledge to improve their comprehension. Typical children used the main idea to make causal and sequential connections to the
text, which improved their comprehension predictably. The autism group did not make these connections, demonstrating poorer comprehension skills overall (Nuske & Baven, 2010).

Children with autism are known to have difficulty getting the main idea or the ‘gist.’ This autism symptom is often referred to in a theory called, ‘The Weak Central Coherence Account’ (Diehl et al., 2006; Nuske & Baven, 2010). Rather than seeing objects, or even people, as a ‘whole,’ children with autism tend to focus on details (Naples, 2010; Nation, Clark, Wright, & Williams, 2006). Recent research refers to this extreme focus on details as ‘finer grain size processing’ (Naples, 2010). This means that children with autism may view the world in smaller, more precise units, rather than as a coherent whole.

Literature supports the idea that children with weak mental imagery also tend to struggle to get the main idea in a story (Bell, 1991b; Oakhill & Patel, 1991). Researchers have found that children with weak imagery struggle with forming, “an integrated and coherent model of what they have read or heard” (Oakhill & Patel, 1991). Children who were taught, in previous studies, to use imagery could not only grasp the main idea but use that ‘gestalt’ to comprehend the whole reading passage (Joffe, Cain, & Maric, 2007; Bell, 1991b; Oakhill & Patel, 1991). It is possible that the reason why children with autism struggle with the main idea is that they have impaired mental imagery. Perhaps, the previously discovered autism feature of ‘weak central coherence’ (Nuske & Baven, 2010) has its basis in a cognitive difference in the use of mental imagery. This theory will be empirically tested in the study that follows.
Integration & the Functional Under-Connectivity theory

Integration is another important skill that has been shown in the literature to be related to imagery ability. That is, children with weaker imagery tend to remember and provide less information about how story events are integrated and connected (Cain, 2009; Oakhill & Patel, 1991). Many experts in comprehension related imagery skills model imagery as a constructive process whereby the reader integrates across story elements, making key inferences, and creating a coherent picture of the text (Cain, 2009; Gambrell & Bales, 1986). These images then provide a scaffold or structure on which story comprehension rests. Poor imagers are unable to integrate all of their images, and much of the meaning is lost.

A body of literature exists to support that even though children with autism can remember as many facts as typical children, these facts are not connected and integrated well when they retell stories (Deihl, Benetto, & Young, 2006; Heumer & Mann, 2010; Nuske & Baven, 2010; Tager-Flusberg, 2006). There is also a great deal of research supporting the notion that children with autism do not integrate details or understand conversational context (Nuske & Baven, 2010; Walberg, 2001; White, Happe, Hill & Frith, 2009). They fail to integrate background knowledge with new learning or to integrate information from one part of the story with information from another part (Norbury & Bishop, 2002; Nuske, 2010; O’Connor & Klein, 2004; Walberg, 2001). Further, children with autism tend to focus on irrelevant details, rather than gathering a coherent mental representation of what they read (Heumer & Mann, 2010; Nation, Clarke, Wright, & Williams, 2006; Nuske & Baven, 2010). This conclusion that children
with autism fail to ‘mentally represent’ the story as a ‘coherent whole’ is consistent across studies.

While imagery skills in the autism population have scarcely been considered, there is evidence for the idea that the autism brain does not integrate images and words effectively. Kana, Keller, Cherkassky, Minshew, and Just (2006) studied the autism brain during imagery related tasks. Twelve individuals with autism (mean age = 22, SD = 8 years) were compared to 13 typical individuals, matched for age and IQ. In the experimental condition, participants were presented high and low imagery sentences on a computer screen, and for each, they answered questions as to whether the sentences were true or false. As the sentences were displayed, a functional MRI was conducted. Kana et al. found that people with autism showed more activation in comparison to controls in the parietal and occipital regions (responsible for visual images). However, the areas of the brain that process language, were not well synchronized with the parts of the brain that process imagery. Thus, the autism brain seems to show ‘functional under-connectivity’ (Kana et al., 2006).

Just, Cherkassky, Keller, and Minshew (2004) also examined brain activation in autism during the presentation of sentences. Participants in the study included 17 children with autism and 17 neurotypical children in the control group. The ages of participants were not given in the study report; however, the groups were statistically matched, on the basis of age, IQ, gender, and family SES. Sentences were presented and then participants responded to simple probes. The two groups were compared using a functional MRI during the sentence tasks.
Just et al. (2004) found a significant difference in the autism and typical groups in terms of brain functioning. The autism group showed a lack of functional connectivity between the two regions of the brain that are responsible for language comprehension: Broca’s area and Wernicke’s area. The authors report,

Note that our analysis shows that the activation between two areas is less synchronized in the autistic group specifically at the time that they are doing the sentence comprehension. (p.1816)

Thus, overall, it seems that the autism brain is not well integrated or functionally connected, leading to a lessened ability to use imagery effectively.

Further extending this research, Just (in press) examined the functional connectivity of the autism brain during a functional MRI in a group of adults with autism as compared to matched controls. Just and colleagues claim that they were studying, “two somewhat separable neural systems, the mental imagery and the language processing systems” (p.2). The study includes pictures of the brains of the control group as compared to the autism group. There is a striking difference in the pictures. The autism group shows activity in the visual centers regardless of the level of imagery required for the task. The authors claim that this difference is consistent with the theory that people with autism prefer visual strategies, making reference to Temple Grandin’s 1995 book, *Thinking in Pictures*.

Interestingly, these strengths in visual-perceptual abilities are over-utilized and under-integrated in autism (Just, in press; Just et al., 2004, Kana et al., 2004). According to this research, people with autism seem to rely on visual strategies, even when the task
does not require them. Further, even though the visual centers are active, it is clear from the MRI pictures that there is a lack of integration with other areas of the brain (Just, 2004; Just in press; Kana et al., 2006). The control group’s brain picture displays just two main active brain centers but there is activation across the two (Just, in press). The activation levels in the autism brain pictures appear like a random smattering of polka-dots (Just, in press). Thus, this research indicates that the brain centers do not ‘talk to each other’ effectively in autism (Just, in press; Just et al., 2004, Kana et al., 2006). This study replicated and corroborated the findings of past research indicating an under-connectivity between the regions of the brain responsible for imagery and language in autism.

All of this brain research on imagery in autism points to the same basic conclusion: the autism brain seems to be under-connected (Just et al., 2006; Kana et al., 2006). Even though people with autism tend to prefer visual strategies, they over-use these strategies, even when irrelevant to the task (Just, in press). Further, the imagery centers are not communicating with the language centers, which makes comprehension difficult. Thus, theory suggests that children with autism may have specific imagery deficits but there is not currently a measure to assess integrative-comprehension imagery in autism.

**Imagery Assessment**

The assessment of imagery is not a new concept. Imagery assessments have been in circulation since the year 1880 with the invention of the Breakfast Table Questionnaire (Galton: 1880). Galton was interested in the impact of gender and age on imagery; as
well as, whether or not imagery skills could predict human behavior. While the instrument did not predict behavior well (Hall, Pongrac & Buckholz, 1985), it has been widely used and validated as an imagery measure. This instrument was expanded with the invention of a 150 item questionnaire called the Questionnaire Upon Mental Imagery (QMI: Betts, 1909). Even with the early date when this instrument was created, the majority of studies that include an imagery assessment rely on some form of the QMI (Cook, Melamed, Cuthbert, McNeil & Lang, 1988; Danaher & Thoresen, 1971; Hall, Pongrac, & Buckholz, 1985; Hiscock, 1978; Isaac & Marks, 1994; Pitman, Lasko, & Herz, 1993).

The most current version of the QMI is a shorter form of the instrument, rated on a 7 point scale for vividness; which was demonstrated to show a strong correlation with Bett’s 1909 version (Hiscock, 1978). Building on this research, an imagery scale was developed and validated called the Vividness of Visual Imagery Questionnaire (VVIQ: Marks, 1973). Isaac and Marks (1994) effectively demonstrated the internal consistency and construct validity of the QMI and VVIQ instruments in a study of over 655 participants, from age 7 to over 50+ (Isaac & Marks, 1994).

Isaac and Mark’s intent was not to validate the instruments. Rather, this study was an analysis of occupation, age, and gender differences in self-reported imagery usage. The researchers acknowledge that questionnaire measures have received a great deal of criticism; but conclude that: the significant correlation between results on the two instruments, the internal consistency of scores, and construct validity demonstrated in
such a large sample, is unlikely to have occurred by chance. Their findings indicate that creation of a valid and reliable imagery assessment instrument is possible.

Another imagery instrument was developed to test motor imagery called the Imagery Exercise Questionnaire. Hall et al. (1985) argued that much of the imagery research had been restricted due to inadequate measures of imagery. Hall and colleagues developed a new instrument to measure motor imagery. This instrument assessed the ability to perform visual rehearsal prior to performing an athletic task. The Movement Imagery Questionnaire (MIQ) was validated on a sample of 74 students enrolled in physical education classes and the psychometric properties were demonstrated (Hall et al., 1985).

Imagery assessments have also been conducted in the areas of learning and memory. Children are taught to use imagery to remember a list or to associate the paired objects (Pressley, 1996, 1977). The objects were language-based but not meaningfully associated. The child may be asked to pair a car with a hammer, for example. The theory was that imagery would aide in their memory for these associations based on strong research that memory for concrete objects improves with the use of imagery strategies. The researchers assessed participants by simply presenting the objects; testing their memory for the objects; reminding them to use imagery to remember the objects, and then testing them again. Results indicate that memory improves when imagery is used (Clark & Paivio, 1987; Pressley, 1977; Pressley & Levin, 1977). A similar approach was used by other researchers in the area of learning. Gambrell and Bales (1986) read imagery-rich passages; and instructed examinees to ‘make movies when they read.’
Results from pre and post tests were that children demonstrated memory gains when they used imagery (Gambrell & Bales, 1986; Joffe, Cain & Maric, 2007).

A review of the literature reveals only one observation measure for imagery, the Visualizer/Verbalizer Behavior Observation Scale (VV-BOS: Luetner & Plass, 1998). This instrument was designed to measure the learning style preferences of college students. In this assessment, students are set up in a naturalistic learning setting as part of a class in German vocabulary. The students are tested for their learning style preferences by a computer program which presents them with verbal and visual options. Students are asked whether they would like to have the vocabulary word described to them through images or pictures. Results of this analysis were intended to reveal whether people preferred either the verbal or visual learning modality.

This is certainly an innovative instrument for assessing learning style preferences. However, this scale would not be helpful in a study like the one here. Research suggests that children with autism do indeed prefer the visual learning modality (Grandin, 1995; Just, in press). The problem in autism is not an issue of preference but one of integration. As discussed previously, there seems to be a functional under-connectivity in the areas of the autism brain that process images and language.

A review of the literature revealed one unique assessment of imagery for comprehension (Sadoski, 1983). This assessment is more of an unstandardized-checklist approach but it is the first of its kind. Sadoski assessed imagery and considered the impact that these imagery skills have on standardized comprehension tests. In this study, 48 students in the 5th grade in a Connecticut public school completed two standardized
comprehension tasks, and an unstandardized story-retell task. The imagery question occurred immediately following the story-retell task. The researchers recorded the number of reported images in addition to data about whether or not they remembered an image of the climax of the story.

There was a general trend indicating that the children who remembered an image of the story’s climax performed better on comprehension tests (Sadoski, 1983). However, the number of images recalled did not impact comprehension predictably. Sadoski hypothesized that comprehension tests rely almost exclusively on verbal questions; not adequately measuring the relationship between imagery and language abilities. It is possible that children who reported more images may have performed better on comprehension tasks if the tests had been less verbally loaded. He summarizes by saying that, “The convergence of theories from various perspectives that are relevant to these findings invites further study in this area” (p.121).

Although not an imagery assessment, Bell (1991b) developed an innovative approach to teach and analyze a child’s imagery. Bell (1991b) developed a list of ‘structure words,’ which are provided to help the interventionist evaluate the quality of a child’s images. Further, the structure words could be used to help the child elaborate on his or her images during the intervention. For example, the examiner would ask, “is there any movement in your picture?” or “are there any shapes in your picture?” These structure words were indeed included as potential areas to assess in the early versions of the instrument used in this study. However, Bell’s approach was not intended to be a
standardized imagery assessment; rather, a tool for interventionists to use in order to help children elaborate and improve on their imagery.

Limitations of existing instruments. There are limitations of all these instruments for the purposes of the current study at hand. First, regarding the VVIQ and the QMI, which have the strongest research support, there are standardization concerns about using a 35 year old instrument on a modern-day sample. Further, neither of these instruments was designed for children. These instruments were originally designed as a self-report measure of imagery usage in adults. Further, the VVIQ and the QMI instruments were not designed to assess comprehension related imagery skills.

The approach employed by Sadoski (1983) is an impressive effort to measure imagery skills in relation to comprehension; however, the question of the integration of verbal and imagery abilities was left on the table. Sadoski explained that it was hypothesized, but not yet determined, whether the children performed better when they are able to use a combination of imagery and language strategies. He admits, “this exploratory study allows for few absolute conclusions” (p.120). The exploratory nature of this analysis begs the question as to whether or not the current study can measure integrative-comprehension imagery more precisely. Further, this was a checklist approach, rather than a validated instrument that could be used in this study. Most importantly, none of the instruments mentioned were tested on children with special needs such as autism. Thus, a new scale is necessitated, in order to assess integrative-comprehension imagery in children with autism.
A Model of Integrative-Comprehension Imagery

The model shown in Figure 2 has been developed based on the literature regarding imagery and comprehension. This model will then provide the theoretical foundations for the domains and items to be used in the imagery assessment developed in this study. The branches represent the domains of integrative-comprehension imagery. The roots of the tree are skills that are thought to be the foundations for comprehension, and requiring imagery, but occurring beneath the surface. The trunk of the tree involves in-text connections, sometimes called on-line processing because it refers to basic understanding of the explicitly stated information in the text. Finally, the leaves of the tree are the observable or ‘assess-able’ skills. Most of the domains and leaves will be assessed in this study. A detailed account of each skill is provided just after Figure 3 in the ‘theoretical foundations’ section.
Figure 3

A Model of Integrative-Comprehension Imagery

Marcy Willard, 2012. All Rights Reserved.
**Theoretical foundations of imagery assessment**

**Branches:** Past research has identified the major skills that can be identified through a child’s narrative, and may be impaired in autism, finding two factors: Main Idea and Details (Nuske, 2010). Other research has uncovered two potentially important impaired narrative production skills in the autism population: Main Idea and Inferences (Norbury & Bishop, 2002). Sadoski (1983) used exploratory factor analysis method to identify the imagery domains. He found four potential domains of imagery: Inferences; Central Theme (similar to the Main Idea domain); Mid-level verbal comprehension (Semantics); and a contrast between character details vs. constructivist process (similar to the Integration domain and Details constructs). Oakhill and Patel (1991) identified two major areas of imagery: Integration and Main idea. Others have identified memory for Details as a major factor in imagery (Joffe et al., 2007; Pressley, 1977; Pressley & Levin, 1976). Still others identified Integration and Narrative production (incorporating main idea) as domains. Thus, the 3 major domains used in this study, based on a preponderance of literate, were: Main Idea, Integration, and Details.

**Smaller Branches:** Cohesive representation is the ability to take on a constructivist process while reading, using connections in the story to piece together a coherent, holistic picture (Bell, 1991b; Cain, Gambrell & Bales, 1986; Kinstch, 1988; Norbury & Bishop, 2002; Oakhill & Bryant, 2004; Pearson, in press). Most comprehension experts refer to coherence as an integrative and constructivist process; requiring that the child has developed a mental model of the meaning of the text while simultaneously making text-level connections and factual conclusions (Cain, Oakhill, &
While coherence is not easily assessed, there is some evidence that coherence is observable through a child’s narratives (Cain, 2009; Joffe, Cain & Maric, 2007; Kintsch, 1988). That is, most researchers have found that a child who struggles with forming coherent mental images also tells stories that are not globally coherent (Cain, 2009; Joffe, Cain, & Maric, 2007). Essentially, poor imagers are poor story-tellers. Particularly relevant in this study is the finding that children with autism tell stories that lack clarity and coherence. They provide disjointed narratives that are out-of-sequence and laden with irrelevant details (Nuske & Baven, 2010; Snowling, 1986; Wahlberg, 2001; White, Hill, Happe, & Frith, 2009). They often fail to use gestures which would enhance the listener’s understanding of the narrative, and do not mention the emotions of the characters. All of this literature led the researcher on this study to consider ‘Global Coherence’ as another minor domain of imagery assessment (see Imagery Scale Development for full description).

Another minor branch that was included in this model is the ‘linking connections’ branch. The items within ‘connections’ include skills such as making causal and sequential connections; and ‘time links.’ Time links refers to the temporal order or the time of day of the events. A child demonstrating this skill might say, “oh, and now it got dark out; must be night time.” The items of ‘causal connections’, ‘time of day’ and ‘sequencing’ were all included in the imagery instrument. Some of these items were
deleted in later versions of the scale due to reliability or variability concerns which will be explained later (see Imagery Scale Development).

**Roots:** Research has identified inference-making, also known as ‘gap filling’ (Briton & Gulgoz, 1991; Kintsch, 1994); as an important skill in of comprehension and imagery (Oakhill & Patel, 1991; White, et al., 2009). It is ability to make a reasonable guess about what is unknown in the story, using information that was explicitly stated. Of particular interest in this study, children with autism often fail to make inferences (Norbury & Bishop, 2002; Wahlberg, 2001). Further, research has found that when a child with autism makes an inferencing error, he or she holds onto that idea for too long and loses the essence of the meaning (Naples, 2010).

Another ‘root’ skill of integrative-comprehension imagery is ‘comprehension monitoring’ (Cain, 2009; Gambrell & Bales, 1986; Strickland & Feeky, 1985). This is probably the most difficult skill to assess because it involves a coordination of simultaneous and constructivist processes. The concepts referred to above about ‘repairing errors’ are inherent to comprehension monitoring. Comprehension monitoring involves pondering questions like, “Do I understand this?” A child with good integrative comprehension skills uses these self-assessments to adjust his or her thinking; repairing errors, flipping back to check for missed concepts, and researching any unknown vocabulary words.

The final skill considered as a root foundation of integrative-comprehension is ‘linking to background knowledge.’ A wide body of literature has considered the importance of background knowledge to comprehension (Kintsch, 1994; Pearson, in
press; Rapp & van der Broek, 2005). The reason this skill is considered a ‘root’ skill is that research has shown the level of comprehension depends on background knowledge and the reader’s ability to integrate that knowledge with the text (Kintsch, 1994). Evidence suggests that it is not only important to make these links to background knowledge but the links need to be accurate and relevant (Strickland & Feeky, 1985).

Interestingly, in this study, a qualitative evaluation of these data indicates that the children with Asperger’s tended to make many more irrelevant links to background knowledge (See Imagery Scale Development). They elaborated on the story using background knowledge but this actually reduced the quality of their narratives because the information was irrelevant to the story. Thus, although not directly assessable in this study, there was some qualitative evidence of impairment of this comprehension skill in the ASD sample (see Chapter 5: Assessment Results).

The ‘root’ skills are modeled here but could not be assessed in the study that follows directly. These skills are included because researchers are aware that the skills are occurring beneath the surface but they are difficult to observe. As such, the skills mentioned in the ‘leaves’ of the model will be directly assessed; the roots are theoretically foundational to these assessed skills.

The trunk: The trunk of the tree is the on-line processing or text-based comprehension. ‘Making connections within text structure’ is a skill belonging to the trunk of the tree because of the level of cognitive processing. Some of the deeper level processes occur at the roots and are more subconscious and simultaneous (inference-making, monitoring, and linking to background knowledge). However, text-level
processes are sequential and generally more overt or observable (Cain, Oakhill, & Bryant, 2004). The reader is not expected to infer information during this process; text-level comprehension is simply a matter of piecing together what is known in the story. Also, within the trunk are cognitive processes like ‘dual coding’ which involves the integration of images and words. Semantic categorization is a text-level process (and also a component skill of cohesive representation); referring to the ability to group words into meaningful categories based on common features (Mehta, Newcombe & Haan, 1992; Thompson-Schill, Aguire, Esposito, & Farah, 1999). However, semantic text-level processing was not possible to assess in the study that follows because of the limitations of using archival data. In this study, the text-level processes considered were: causal connections and sequencing.

Research shows that children with poor imagery struggle with making causal (Bell, 1991b; Cain & Oakhill, 2004) and sequential connections (Clark & Paivio, 1987). Similarly, children with autism tend to fail to make these connections, potentially due to impaired executive functions (Rutherford et al., 2007). Children with a poor understanding of cause-and-effect relationships are impaired in a variety of social skills, planning tasks, and comprehension overall. Sequential connections require readers to put their mental images in order, based on the plot-line of the story. For example, “First he woke up, then brushed his teeth, and then left for work.” Children who struggle with sequencing also have grave difficulties with following directions (Bell, 1991b).

Although, sequencing and causal connections skills were modeled under ‘Linking
Connections,’ they were later captured under the Integration domain based on panel review and factor analysis (See Chapter 4: Imagery Scale Development).

_The leaves:_ The leaves of the model are the more readily observable imagery-related comprehension skills. While the leaves are shown as connected with the branches, there is considerable debate in the literature about which component skills fall under the major domains. The decisions about where to capture the skills in each domain was made through expert panel review and factor analysis (See Scale Development). Each of the leaves selected for the imagery assessment was chosen because the skill is amenable to assessment through observation. Pressley (1976) considered allowing the children to draw their images to assess imagery, and have found this approach ‘irrelevant’ (Pressley, 1976). Researchers have suggested that it is critical to teach directly to the task; which is comprehension (Pressley, 1976, 1977); rather than imagery, which is the strategy being used to complete the task (Bell & Bonetti, 2006). Thus, the items chosen for the scale, modeled as ‘leaves,’ were selected because they could be assessed through observation (which was the assessment modality used in this study).

_Main idea leaves._ Children who understand the main idea of stories are able to report on characters, actions, and to recall and describe the major events. In the study here, the Main Idea concept includes an understanding of ‘what happened’ which is symbolized by the ‘actions’ leaf here (Strickland & Feeky, 1985). The actor-as-agent concept is important to the imagery literature (Sadoski, 1983) and the autism literature (Rutherford, Young, Hepburn & Rogers, 2007). Actor agency is the concept that characters are actors on their environments. A child is demonstrating an understanding of
agency when he reports, “the man ate a sandwich” or “the dogs scared the birds away.”

The characters item has support in the imagery literature (Sadoski, 1983) and the autism literature (Rutherford et al., 2007). The characters item is an assessment of the child’s ability to describe the human and animal characters in a story. The sequence item has strong support in the imagery and integrative-comprehension literature (Bell, 1991b; Clark & Paivio, 1987; Strickland & Feeky, 1985). Studies indicate that children with autism show poor sequencing skills (Rutherford, et al., 2007). This item falls between the Main Idea and the Connections or Integration described previously. The decision to move sequence to Main Idea domain will be discussed in Chapter 4: Scale Development.

Integration leaves. In this study, the Integration branch includes the ability to make integrative statements (Cain, 2009; Clark & Paivio, 1987, Oakhill & Patel, 1991), to take on the perspective of the characters (White et al., 2009), to use gestures (Bell & Bonetti, 2006), and to form integration models (Pearson, in press).

Gesture use has a known basis in imagery. Imagery experts Bell and Bonetti explain, “Gesturing represents imagery and enhances verbal descriptions” (p.27). They hypothesize that children who do not use gestures are likely to also show impaired imagery. As such, gestures are included as a leaf in this model and an item on the imagery assessment that follows in Chapter 5: Imagery Scale Development. It should be stated that the gestures item was developed on a sample of mostly American, White, Non-Hispanic participants. Thus, this item is based on mainstream American gestures; however, the item would need to be adapted in order to apply it to other cultures.
The integrative statements leaf represents comments provided by the child indicate that he or she connecting ideas throughout the story or relating personally to the narrative. For example, a child might say, “he’s eating at bed-time. I wonder if he will have a tummy-ache. That happened to me once…” Perspective taking is the ability to understand what he character’s goals and intentions; thoughts, and feelings (Sadoski, 1983; Strickland & Feeky, 1985). Thus, the Integration branch is thought to include integrative statements, gestures, and perspective taking. Each of these skills will be assessed in the Imagery Scale Development section (Chapter 5).

**Details leaves.** Finally, the leaves of the Details domain included: gathering factual knowledge and recalling specific details. The Details might include the setting of the story, the number of characters, or the size of objects in the story (Cain, 2009; Pressley 1976; Strickland, 1985). Children with stronger comprehension and imagery can remember more details than children with weaker imagery. Researchers, Joffee, Cain, and Maric (2007), showed that children with specific language impairments (SLI) who were taught to use imagery could recall more information about the literal facts and important details than children who did not use imagery effectively. Thus, this research demonstrates: a) that imagery can be taught, and b) that Details should be considered in an imagery assessment.

Sensory details are highly associated with imagery. Bell & Bonetti (2006) explain, “Imagery is a sensory-cognitive function basic to many kinds of language processing” (p.5). Bell & Bonetti believe that one of the primary goals of imagery intervention is to, “bring sensory information imagery to a conscious level” (p.5). In this
study, sensory included elements such as: taste, smell, color, shape, and sounds. These items were included as ‘leaves’ in this model and items on early versions of the WIOS (See instrument development).

Taken together, this model of integrative-comprehension imagery includes three major domains and several associated items within each domain. These component skills found in the literature will be used in this study in the development of a proprietary imagery assessment scale. Further, these items were carefully selected not only for their underpinnings in comprehension and imagery; but also, because these are the precise skills suspected to be impaired in autism. Thus, the researcher wanted to know whether or not children with autism are impaired on the skills that are known in the literature to require imagery. The research questions were as follows.

**Research questions**

1) Can a valid and reliable scale be developed to assess integrative-comprehension imagery skills in children with and without autism?

2) Do children with autism show different abilities than typically developing children in terms of imagery-related integrative-comprehension skills as assessed by the imagery scale?

   a. Do children with autism show more difficulty in understanding the main idea of a story as measured by the Main Idea domain of the imagery scale? (the Weak Central Coherence Theory)
b. Do children with autism show more difficulty integrating story events and ideas as measured by the Integration domain of the imagery scale? (the Functional Under-connectivity Theory)
Chapter Three: Method

This study is a secondary analysis of data collected by a research team at JFK Partners, Center of Excellence in Autism and Neurodevelopmental Disabilities in partnership with the University of Colorado Medical School and The Children’s Hospital of Denver. The participants were part of a longitudinal study of the development of children with autism and other developmental disorders, which was originally funded by the National Institutes of Health, entitled: A Longitudinal Study of the Developing Phenotype of Autism (COMIRB Protocol Number: 96-587). The study utilized a full battery of tests assessing: diagnosis, intelligence, language, psychological symptoms, motor skills, and global adaptive skills. All subjects in the original study signed informed consent forms to participate, to be videotaped completing assessments, and to allow researchers to analyze these data for research. Dr. Susan Hepburn, the Principal Investigator on this study, was a research advisor on the current study.

Design

This study employed two major analysis phases; employing two different research designs. First, the instrument validation process was designed based Benson and Clark’s (1982) model for instrument development. Benson and Clark’s model includes four phases: I: Planning, II: Construction, III: Quantitative Evaluation, and IV: Validation. A similar model to this is seen in many instrument development dissertations and textbooks (Bahraini, 2008; DeVellis, 2003).
Secondly, a multivariate analysis was conducted on the imagery assessment results to determine whether or not diagnosis was a significant predictor of imagery scores on the WIOS, controlling for the effects of IQ, Age, and Gender. This analysis utilized an ex-post-facto design; also known as causal-comparative design (Johnson & Christenson, 2008). As the name implies, this design involves ‘after-the-fact’ analysis of previous events. The design is non-experimental because no manipulation of the independent variables occurred. Further, random-assignment was not possible in this study because participants were grouped based on diagnosis (Johnson & Christenson, 2008).

The primary limitation of causal-comparative design is that the conditions necessary for causality cannot be met: 1) temporal precedence, 2) correlation, and 3) isolation (‘the third variable problem’) (Johnson & Christenson, 2008). In this study, the requirement of temporal precedence was not possible to achieve because data were all collected at the same time point. However, in order to strengthen the predictive power of this design, the other two conditions for causality were addressed.

To meet condition #2; correlation, a relationship between imagery related-comprehension problems and autism diagnosis was identified in the literature. Correlations were also analyzed with regard to whether the items on the WIOS correlated with diagnostic status (See Chapter 5: Imagery Assessment Results). Condition #3; isolation, was met. Variables such language ability, symptom severity, and cognitive ability were controlled for through inclusion criteria. Demographic variables such as socio-economic status, parental education, and ethnicity were consistent across the
groups. Finally, variables like IQ, Age, and Gender were included as statistical controls. Thus, although the requirement of temporal precedence could not be met; the predictive power of the design was increased because the causality conditions of correlation and isolation were addressed.

**Power Analysis**

Prior to conducting the study, power analysis was calculated in order to determine the number of participants necessary in order to find statistical significance, given a true effect exists in the population (to avoid a Type II error). Power analysis was calculated using G*Power 2. This software is known to provide precise power tests for a certain anticipated effect size and alpha level. Power levels of over .70 are generally considered acceptable (Stevens, 2007). The study is, thus, deemed worthwhile because there is a 70% chance of finding a significant result (Stevens, 2007, p.111). For the Power analysis, a Moderate Effect Size was utilized, an alpha level of .05, a Power level of .70, utilizing 4 predictors. Results indicated that the number of participants required was N=70 (Fcrit=2.513). Therefore, a study comprised of roughly 70 participants is likely to find statistical significance, if indeed there is a true effect.

**Participants**

There were 212 participants in the longitudinal archival data set, ranging from Time 1 (toddler), to Time 2 (preschool), and Time 3 (school-age). By Time 3 of this study, from which data for this study were harvested, 107 participants remained, with a wide range of diagnoses (autism, Asperger’s, PDD-NOS, Developmentally Disabled, Fragile X, Down syndrome, Typical, and others). Time 3 data included children who
ranged in age from 5-13. The children were similar in terms of demographic variables such as: race and ethnicity, socio-economic status, and mother’s education level. The children were not well matched in terms of psychological symptom severity, intelligence, age, and language ability. It was thus determined that a rigorous selection process must ensue in order to narrow down the sample to the participants that were similar enough on these covariates in order to be effectively compared in a causal-comparative design.

**Inclusion Criteria**

For the current study, data were extracted from Time 3 of the original longitudinal study which included 107 cases. Participants were selected for the final data set based on: Diagnostic Status, IQ, Age, and Language Ability. Thus the final sample, after deleting cases that did not meet inclusion criteria included a total of 71 cases (see Figure 4).

**Figure 4**

*Sample Selection Procedure*
**Inclusion Rule #1: Diagnostic Status.** The umbrella term of autism spectrum disorders (ASD), as described earlier, includes children who share some common features of autism but do not meet full criteria for an autism diagnosis. In the current study, children with autism, Asperger’s Disorder, and PDD-NOS were included. Based on pilot testing, it appeared that performance on the WIOS indeed fell along a similar continuum as does the algorithm used for diagnosis. That is, the Typical group seemed to perform better than the ASD group (including PDD-NOS and Asperger’s); and the ASD group performed better than the autism group. Thus, although the archival data set included children with a variety of developmental disorders, the sample selected for this study included only 3 diagnostic categories: Autism, ASD, or Typically developing.

**Inclusion Rule #2: IQ Score.** Criteria were that cases with a Full Scale IQ under 80 were removed. An IQ score of 70 (Mayes & Calhoun, 2008) 80 (Conti-Ramsden et al., 2006), or 85 (Tager-Flusberg, 2006) was used as an IQ cut-off score in similar studies. The primary investigator on the original longitudinal study suggested a cut-off of 80 (S. Hepburn, personal communications, January 25th, 2011). Dr. Hepburn indicated that IQ tests tend to under-represent the cognitive ability of children with autism. Indeed, test developers for widely used intelligence tests agree that IQ tests underestimate intelligence on the autism population (Wechsler, 2006). In order to address this issue, whenever multiple intelligence tests were included in the data set, the highest IQ score obtained was used in this study. An IQ score cut off of ≥ 80 was used. In the final sample, the range of IQ scores was 80-142.
**Inclusion Rule #3: Age.** Inclusion criteria were that only Time 3 participants would be incorporated into the sample. This group was chosen because they were generally school-aged (age 5-13). Imagery literature indicates that age has an effect on imagery and that pre-school aged children do not yet have fully developed imagery skills (Isaac & Marks, 2007; Pressley, 1977; Pressley & Levin, 1978). Further, the goal of this assessment was to understand imagery skills that might be used for reading comprehension, and generally these skills are practiced in school. The final sample included children who ranged in age from 5-13, with a mean age of 8 years. The researcher acknowledges that this is a large age range. In order to account for this variation, statistical controls will be employed by including Age as a covariate in the regression analysis. Further, it is extremely challenging to find large samples of children with autism who have been clinically diagnosed. As such, other studies have included participants of a similar age range (Mayes & Calhoun, 2008: included children aged 6-14; Nation et al., 2006: included participants from 6-15 years of age).

**Inclusion Rule #4: Language Ability.** In order to control for language problems, children who were assessed with an ADOS-Module 3 only, indicating adequate language ability, were retained. Participants who were given an ADOS Module 1 or an ADOS Module 2, indicating language problems, were excluded from analysis. Examiners on the ADOS are clinically trained to listen for a language sample during their interactions with participants in order to see whether or not they meet criteria to be assessed with a Module-3. Therefore the ADOS module provides one measure of language ability. However, the researcher acknowledges that other measures of language, such as the
Clinical Evaluation of Language Fundamentals (CELF: Semel, Wiig, & Secord, 2003) could have provided better evidence for language ability. However, these scores were not available on enough of the participants; and as such, only the ADOS module was used for language measure inclusion criteria. The final sample, thus, included 71 cases; all of whom were assessed with an ADOS Module 3.

**Final Sample Description**

The total number of cases meeting inclusion criteria, based on the above decision rules, was n=71. Based on the previously discussed Power analysis, 70+ cases should be adequate to find a statistically significant result, given that indeed one exists in this population. The groups were similar on demographic variables such as: race, ethnicity, mother’s educational level, and socio-economic status. The most common educational level of the mother was ‘some college.’ There were no participants whose mothers had not graduated from high school and fewer than three participants whose mothers had not been to college. All groups had similar socio-economic status. The demographics of the final sample are included in Table 1.
Table 1

Sample Description

<table>
<thead>
<tr>
<th></th>
<th>Entire Sample</th>
<th>Autism (ASD)</th>
<th>Typical Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M(SD)</td>
<td>M(SD)</td>
<td>M(SD)</td>
</tr>
<tr>
<td>IQ</td>
<td>106 (17)</td>
<td>102 (16)</td>
<td>114 (16)</td>
</tr>
<tr>
<td>SES in $1,000/yr.</td>
<td>53 (9.6)</td>
<td>54 (6)</td>
<td>52 (11)</td>
</tr>
<tr>
<td>Age in years</td>
<td>8 y (2 y)</td>
<td>9 y (2 y)</td>
<td>8 y (1.8 y)</td>
</tr>
<tr>
<td>N (%)</td>
<td></td>
<td>N (%)</td>
<td>N (%)</td>
</tr>
<tr>
<td>Race (Caucasian)</td>
<td>68, 95%</td>
<td>46, 95%</td>
<td>22, 96%</td>
</tr>
<tr>
<td>Gender (Boys)</td>
<td>52, 73%</td>
<td>42, 88%</td>
<td>10, 43%</td>
</tr>
<tr>
<td>Ethnicity (Hispanic)</td>
<td>5, &lt;1%</td>
<td>1, &lt;1%</td>
<td>4, 2%</td>
</tr>
<tr>
<td>N</td>
<td>71</td>
<td>48</td>
<td>23</td>
</tr>
</tbody>
</table>

Overall, there were a total of 71 participants. Of these, 48 of the participants had a diagnosis on the autism spectrum; whereas, 23 of them either were typically developing or had a mild disability during childhood but presented typically by the time of this study. The overall sample showed a mean IQ of 106 (SD=17). The Autism Spectrum group had a mean IQ of 102 (SD=16). The Typical group showed a mean IQ of 114 (SD=16). Thus, the overall sample had a mean IQ in the average range. The Typical group showed a mean IQ of 114, which is 12 points higher, on average, than the Autism group. This difference is certainly important to note; however, it is less than one standard deviation from the mean for the intelligence tests used in this study (SD=15).

In the Autism Group, there were 42 boys and 6 girls. The typical group included 10 boys and 13 girls. This higher population of boys in the autism sample may be due to the fact that boys with autism outnumber girls 5:1 in the general population (CDC, 2012). The demographics for both groups were similar in terms of socio-economic status with annual household incomes in the range of $52,000-$54,000 per year in all groups.

51
groups were similar across race and ethnicity domains, with the vast majority of participants being White, Non-Hispanic. In the overall sample, 2 children were African American and 1 child identified as ‘other.’ The groups were fairly evenly matched for age as well. In the overall sample, the average age was 8 years old. In the autism group, the average age was 9; in the typical group, the average age was 8 years old. Taken together, the groups did not vary significantly on demographic variables, with the exception of gender; where there were significantly more boys in the autism sample.

**Instruments**

**Diagnostic instruments.**

In the original study from whence these data were collected, diagnosis was clinically determined using several valid instruments. The Autism Diagnostic Observation Schedule (Lord: ADOS, 1999) and the Autism Diagnostic Interview-Revised (ADI-R: Rutter, Le Couter, & Lord, 2007) are both tests that were developed for the purpose of differential diagnosis for autism. Researchers consider both the ADOS (Klin, Saulnier, Sparrow, Cicchetti, Volkmar, & Lord, 2007), and the ADI-R to be ‘the gold standard’ for diagnostic assessment of autism spectrum disorders (Matson, Nebel-Schwalm, & Matson). Best practice is to use both instruments in determining the diagnosis (Conti-Ramsden, Simkin, & Botting, 2006). These instruments were designed to align with the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; American Psychiatric Association, 2000). This is the handbook used by psychologists and psychiatrists to diagnose psychiatric and neurological conditions. All of the
instruments utilized in this study show reliability and validity and were administered by clinically trained examiners.

**Autism Diagnostic Observation Schedule**

The Autism Diagnostic Observation Schedule (ADOS: Lord, 1989) is an observation scale whereby a clinician assesses a child’s play and communicative behaviors through a naturalistic observation, using a defined set of developmentally appropriate play activities. The assessment includes approximately 14 activities, 26 items and 3 domains. The reliability and validity of the ADOS has been shown to be strong (Klin, et al., 2007; Matson, et al., 2007) through regular and systematic clinical training sessions where inter-rater reliability and diagnosis verification is assessed. During test development, the instrument was assessed for reliability and found to be adequate, using: test-retest, inter-rater, and internal consistency reliability measures (Matson et al., 2007).

**Reliability:** In this study, the rater agreement was 85% or better. This reliability assessment was conducted over 3 consecutive administrations and was measured on 20% of the cases in the original sample (Philofsky, Fidler & Hepburn 2007; Rutherford, Young, Hepburn & Rogers, 2007).

**Validity:** The test was sent to a large expert panel to evaluate construct validity during test development and found to be strong (Matson et al., 2007). The ADOS was assessed for validity using factor analysis, differential diagnosis, and consistency with DSM-IV criteria during test development (Matson et al., 2007).
**Autism Diagnostic Interview-Revised**

The Autism Diagnostic Interview-Revised (ADI-R: Rutter, Le Couter, & Lord, 2007) is a structured and standardized interview protocol, administered to parents. The instrument includes over 70 items, summarized into an algorithm that includes 3 domains. The ADI-R domains are aligned with DSM-IV criteria.

**Reliability:** For this study, all test administrators were clinically trained for standardization and fidelity and demonstrated over 85% inter-rater reliability with the trainer (Philofsky, Fidler & Hepburn, 2007). Reliability was assessed on 3 consecutive administrations and conducted on 20% of cases.

**Validity:** The ADI-R was tested on 17,173 participants, having autism, autism spectrum, or language disorders (Cox et al., 1999). A thorough review of the autism diagnostic assessments found the ADI-R to be the most valid instrument due to the fact that it was tested on the largest and most varied sample, has more published data regarding psychometric properties, and most closely resembles the DSM-IV criteria (Matson, et al., 2007). Further, longitudinal studies have demonstrated that the diagnostic decisions made with this instrument tend to hold up over time (Cox et al., 1999).

**Intelligence tests.**

There were several standardized intelligence tests included in the study. The children in the sample were generally either given a Wechsler Intelligence Scales for Children, 4th Edition (WISC-IV), or a Wechsler Abbreviated Scales of Intelligence (WASI). Some of the participants were tested with the Differential Ability Scales (DAS), or a Leiter (a non-verbal intelligence test).
Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV)

The Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV: Wechsler, 2003) is a measure of cognitive ability for children aged 6-16 years. Scores on the WISC-IV are measured on a standard scale with a mean of 100 and a standard deviation of 15. The WISC-IV is constructed of 15 subtests, summarized into four composites: Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed. All of the composites combine to form the Full Scale Intelligence Quotient (FSIQ), which is an overall measure of intellectual ability.

Reliability: The WISC-IV is known the literature to have outstanding reliability; ranging from .96-.97 for the Full Scale IQ, and .81-.95 for the other composites (Sattler, 2008). The test was assessed using test-retest reliability (.84-.93) and internal consistency reliability measures (r=.68-.89).

Validity: The WISC-IV was standardized on a sample of 2,200 children from 11 different age-groups, including an equal number of boys and girls, and a wide variety of ethnicities. The criterion related validity results were assessed using other Wechsler tests such as the: WISC-III, WPPSI-III, WAIS-III, and WASI. Although the test shows adequate validity as correlated with these measures (.73-.87 on average), evidence for validity would have been stronger if the test was compared to other cognitive tests from a different publisher (Sattler, 2008). Factor analysis was conducted for construct validity, and results indicate that WISC-IV is a valid measure of intelligence (Sattler, 2008).
The Wechsler Abbreviated Scale of Intelligence (WASI)

The Wechsler Abbreviated Scale of Intelligence (WASI: Wechsler, 1999) is an abbreviated version of the WISC-IV. Similarly to the WISC-IV, the WASI is an assessment of cognitive ability in children aged 6 years to 16 years, 11 months. The WASI provides a Full Scale Intelligence Quotient score which is derived from four subtests.

Reliability and Validity: In order to assess the validity and reliability of the instrument, test developers used a method called ‘counterbalance’ whereby one half of the sample is administered the WISC-IV and the other half is assessed with the WASI, in order to see if the results are highly correlated. This analysis revealed correlations ranging between .78-.84 (Sattler, 2008). The Full Scale IQ scores were compared on the two instruments, finding that children score slightly lower on the WISC-IV than the WASI (an average of 1.8-3.4 points). This means that the tests are correlated but not interchangeable (Sattler, 2008). Overall, however, the WASI is considered in the literature to show acceptable convergent reliability when compared to the WISC-IV (Philofsky, Fidler & Hepburn, 2007).


The DAS-II is a standardized assessment of cognitive ability for children aged 2 years 6 months to 17 years, 11 months (Elliott, 1990, 1995). This instrument is most commonly administered to children from about age 2 to 6 (the early form). The DAS-II was designed to uncover strengths and weaknesses, using a cognitive profile approach. The school-age battery includes 6 subtests; whereas, the early form includes 4 subtests.
The domains are in accordance with the Cattel-Horn Carrol (CHC) model of intelligence. Composite scores are standardized on a scale with a mean of 100 and a standard deviation of 15.

**Reliability:** The reliability of the DAS-II was assessed through test-retest stability and assessed to be adequate; .80-.90. Inter-rater reliability was assessed to be ‘very high;’ (.98-.99) (Elliott, 1995).

**Validity:** The validity of the DAS-II was assessed through confirmatory-factor analysis, to determine whether or not the 11 subtests load on the theoretical factors identified in the CHC theory, and found to be adequate (Elliott, 1995). The test is also correlated with other measures of intelligence; r=.80 (Sattler, 2008). Test developers assessed the predictive validity of the instrument through correlation with achievement tests and it was found to be adequate as well; r=.81 (Sattler, 2008). The standardization sample included children with disabilities, children who had limited English proficiency, and who were deaf (using the American Sign Language version).

**The Leiter International Performance Scale – Revised (Leiter-R).**

The Leiter International Performance Scale, Revised (Leiter-R: Roid & Miller, 1997) is a non-verbal intelligence scale that is used for children aged to 2-21. Rather than requiring verbal instructions or responses, the test makes use of visuals, manipulatives, and other non-verbal stimuli. The domains are: reasoning, visualization, attention, and memory. The Leiter includes 20 tests, which were developed in accordance with the Cattel-Horn-Carrol model of intelligence.
**Reliability:** Reliability for the Leiter was assessed using internal-consistency reliability by age group (Naglieri & Goldstein, 2009). Reliability ranged from .74-.93 on average. Internal consistency reliability was assessed to be .73-.95 which is considered acceptable (Naglieri & Goldstein, 2009).

**Validity:** The Leiter-R was standardized on a sample of 1,719 individuals aged 2-20 (Naglieri & Goldstein, 2009). Validity was tested using measures of fairness, criterion-related validity methods, and confirmatory factor analysis; and was found to be adequate (Naglieri & Goldstein, 2009). The test correlates with the Stanford-Binet-Fourth Edition (another standardized measure of intelligence) and the WISC-IV with correlations ranging from .55-.75. The Leiter-R has been shown to be a fair assessment of children from a large variety of clinical populations and ethnic groups (Naglieri & Goldstein, 2009).

**Outcome Instruments**

**ADOS, Module 3: item 6.**

The Autism Diagnostic Observation Schedule (ADOS: Lord, Rutter, DiLavore, & Risi, 1999) is an individually administered naturalistic assessment of play, social, and communication skills that are considered to be important in consideration of an autism diagnosis. The ADOS-Module 3 was selected because examiners only administer the Module 3 to examinees who have adequate language ability.

The ADOS-Module 3 includes 14 activities or ‘tasks.’ Of these, the current study only evaluated responses on the ‘Telling a Story from a Book’ task, which is item 6 on the scale. This item was chosen because the task requires children to tell their own
narratives. The task is also standardized such that children read the same story, in the same clinic room, are asked the same questions about the story, and provided with the same prompts.

This study employed a secondary analysis of this task, allowing researchers to consider whether or not children provide imagery-rich discourse when they tell stories. Traditionally, this task is an assessment of autism symptoms with regard to social reciprocity, eye-contact, and general interaction skills. In this study, however, item 6 was used to evaluate imagery. Other studies have employed a similar approach, using archived video-tapes of the ADOS for a language-related assessment (Tager-Flusberg, 2006). It is important to note that the archived videos of the ADOS were used to develop the WIOS; however, the WIOS can be used on any narrative. This particular task was chosen because the children were asked to tell their own stories and because of the standardization. However, it is possible that the WIOS approach could be applied to any story-telling task; either in a school or clinic environment.

In this task, the examinee is shown a story-book with vivid illustrations but few words. Generally, a book called Tuesday (Wiesner, 1991) is used. The examiner begins telling a story about the pictures in the book and then stops. The child is then asked to begin telling his own story. For this study, each child’s videotaped responses to this item were observed in order to assess imagery skills in the participants (see Willard Imagery Observation Scales).
Willard Imagery Observation Scales

The Willard Imagery Observation Scales (WIOS: Willard, 2012) is a narrative observation scale used to assess imagery abilities. The WIOS includes 8 items, summarized into 2 domains: Main Idea and Integration. The scale was developed for this study using a standardized multi-phase approach to instrument development provided by Benson & Clark (1982). The instrument design and development process is detailed in chapter 4: Imagery Scale Development.

Reliability and Validity. Reliability was assessed using internal-consistency reliability (.70) and inter-rater reliability and was found to be good (.85-91). Face validity, construct validity, and concurrent validity were assessed to be adequate (see Chapter 4 for a thorough description).

Overall, there were three types of instruments used in this study: diagnostic instruments, IQ tests, and the Willard Imagery Observation Scales. The instruments used for diagnosis were the ADI-R and the ADOS. For inclusion criteria considerations, various intelligence tests were used: WISC-IV, WASI, DAS-II, and Leiter-R. Videotaped administrations of the ADOS were also used in development of the Willard Imagery Observation Scales (although the scale could be used on any story). Development and validation of the WIOS is discussed in Chapter 4: Imagery Scale Development.

Analysis Plan

The researcher was interested in imagery-related integrative-comprehension skills in children with autism as compared to typically developing children. In order to analyze whether or not diagnosis (Autism, ASD, or Typical) was a significant predictor of
imagery-related item scores, a multiple regression statistical procedure was chosen. In order to assess imagery, the Willard Imagery Observation Scales (WIOS) was developed. Thus, the first part of the analysis plan was the Imagery Scale Development and the second part was the Imagery Assessment Results analysis.

**Imagery Scale Development**

The Willard Imagery Observation Scales was developed in a multiphase approach proposed by Benson & Clark (1982). Phase I, the Planning phase includes:

- operationalizing the construct to be measured and the test domains,
- reviewing literature on the test domains,
- and identifying the target groups. In Phase II, the Test Construction phase; an initial item pool is developed, a qualitative evaluation by judges ensues and scoring criteria are defined. Phase III is the pilot testing phase where the instrument is tested for reliability. Phase IV is the validation phase. The instrument was tested using this four phase model, in order to determine whether or not the instrument showed adequate reliability and validity; such that it can be used in the Imagery Assessment Results analysis that follows.

**Imagery Assessment Results**

Once the scale was validated, it could be used to assess imagery in autism, which was the aim of this study. All of the 71 cases in the sample were assessed, using the Willard Imagery Observation Scales. Participants received a score of 0-2 on each of 8 items, falling into two domains. The scores of the autism sample could then be compared to the typical sample. The statistical procedure used in data analysis was multiple-regression. The variable of interest was Diagnosis; used to predict scores on the imagery
scale. In addition to these variables of interest, a multiple regression allowed for the inclusion of IQ, Age, and Gender as control variables.

**Assumptions.** Before the regression analysis was conducted, it was important to check that all of the major assumptions of regression have been met. The primary assumptions of regression are: independence, homoscedasticity, normality, and error mean of 0. Each of these assumptions were met, and are described in Chapter 5: Imagery Assessment Results.

**Selection of Statistical Control Variables.**

The variable of interest was Diagnosis. The participants were compared based on a diagnosis of autism, ASD, or Typical. The Diagnosis variable included three categories coded thusly: Autism = 1, ASD = 2, and Typical=3. This decision was made because autism spectrum disorders are diagnosed based on a continuum of symptom severity. Thus a ‘1’ was used to represent the most severe symptoms; a ‘2’ was used for less severe symptom profiles, and a ‘3’ indicated no autism symptoms. This approach was carefully vetted in consultation with the principal investigator on the archival data set, and in consultation with an expert in statistics.

However, when direct comparisons are made between children with autism as compared to typically developing, all of the ASD categories were considered part of the ‘autism group.’ These categories were only used to allow for visual inspection of data, not for analytics. This approach was used in describing the sample demographics; because it was important to show how the ASD group compared to the Typical group on important covariates like IQ and Gender. This same approach was used in describing the
score distribution. For example, the scoring results in Chapter 5 include a table of the Mean and Standard deviation of scores on each item. This analysis was grouped by Autism vs. Typical; such that visual inspection would indicate how children with autism symptoms score on the scale as compared to typically developing.

In order to isolate Diagnosis as the predictor variable, it was necessary to consider which other covariates or potentially spurious relationships could confound results. The variables selected were: IQ, Age, and Gender because all of these are known in the literature to bear a relationship to the imagery tasks (Isaac & Marks, 1994).

**Control Variable #1: IQ.** First, IQ could be a significant mediator of imagery-related comprehension skills. It was not possible to match the groups for IQ as the typical group had an average IQ that was 12 points higher as compared to the autism group. Initially, IQ was intended to be controlled for by inclusion criteria by only including cases with an IQ of 80 or greater. However, the average IQ was still disparate among groups; and thus, it was necessary to control for IQ with statistical procedures in addition to inclusion criteria.

**Control Variable #2: Gender.** Next, it was important to consider gender as a mediator of the effects of diagnosis on imagery scores. Evidence in the literature indicates that there are gender differences in imagery (Isaac & Marks, 1994). In the current study, there were a significantly higher percentage of males in the autism sample as compared to the ratio of males to females in the typical sample. This higher percentage of males in the autism sample is common in autism research as males outnumber females with autism 5:1 in the general population (CDC, 2012). These past research findings,
coupled with the disparate numbers of girls and boys in the sample, led to the conclusion that gender should be included as a covariate.

**Control #3: Age.** Finally, the *age* parameter was considered. Variation in age was initially controlled for through inclusion criteria by only including Time 3 data in participant selection. However, there was still a large variation in age of the overall sample, ranging from 5-13 years of age (Mean age = 8). Although, not ideal, a similarly large age range was used in other studies regarding comprehension in autism (Mayes & Calhoun, 2008: included children aged 6-14). Age is known in the literature to have a moderate effect on imagery skills (Isaac & Marks, 1997; Oakhill & Patel, 1991; Pressley, 1977; Pressley & Levin, 1978). In light of the existing literature and the large age range of the sample, age was included as a covariate.

Thus, it was determined that three covariates should be considered: IQ, Age, and Gender. Diagnosis was the predictor variable of interest. Each of these variables was included as independent variables in the regressions. In order to answer the two primary research questions regarding imagery in autism, an imagery scale was developed. Then, that scale was applied to a sample of children with and without autism so that their scores could be compared. The section that follows details the process of instrument development (Chapter 5: Imagery Scale Development). Immediately after, is a multivariate analysis comparing the imagery skills of children with autism to neurotypical controls (Chapter 6: Imagery Assessment Results).
Chapter Four: Imagery Scale Development

The intent of this section is to detail the process used to develop, and assess the psychometric properties of, the Willard Imagery Observation Scales. Approval for this study was obtained from the University of Denver’s Institutional Review Board. The pilot study was approved on 5/23/2011 (DU-IRB Protocol #2011-1782). The full study was approved on 1/27/2012 (DU-IRB Protocol #2011-1975). The researcher acknowledged that validation of the instrument would not be complete during the course of this dissertation study, as the validation process is generally a lengthy, costly, and ongoing endeavor (Benson & Clark, 1982). However, in order to utilize the instrument as a basis for understanding imagery, it was necessary to assess the reliability of the instrument and to begin the validation process. This section will answer Research Question 1: Can a valid and reliable scale be developed to assess integrative-comprehension imagery in children with and without autism?

The method used to assess the reliability and validity of this scale was adapted from the scale development model provided by Benson & Clark (1982). This method includes four phases. Phase I is the Planning Phase; Phase II is the Construction Phase; Phase III is the Reliability Analysis; Phase IV is the Validation phase. The following flow-chart depicts an adapted version of the Benson & Clark model. A similarly adapted version of the model has been used in other studies (Bahraini, 2008).
Figure 5

*Instrument Development Flow Chart* (Adapted from Benson & Clark, 1982)

**Phase I: Planning**
- State test purpose
- Operationalize the construct and define test domains
- Review literature to confirm that test is needed
- Identify target groups

**Phase II: Test Construction**
- Develop initial item pool
- Qualitative Evaluation by Judges
- Define scoring criteria

**Phase III: Reliability**
- Pilot test instrument (N=25)
- Conduct statistical and substantive analysis
- Revise items and test domains
- Test reliability of final scale

**Phase IV: Validation**
- Redistribute instrument to validation sample (N=46)
- Analyze psychometric properties of the scale
- Continued validity tests
Phase I: Planning

- State test purpose
- Operationalize the construct and define test domains
- Review literature to confirm that test is needed
- Identify target groups

State Test Purpose

The first step in the Planning phase was to define the purpose of the test. This purpose statement allows the test developer to ensure that the test is necessary, and not already existing. According to the Benson and Clark (1982) model, the purpose of the test should be written out in a clearly defined statement. The purpose of the WIOS is to score observations of children’s narratives in order to provide an assessment of their abilities to tell stories using imagery-rich stories, indicating strong imagery and integrative-comprehension skills.

As previously provided in the literature review, observation can be a useful means of assessment, and has been used effectively with other imagery measures (Visualizer-Verbalizer Behavior Observation Scale: Leutner & Plass, 1998). In this study, the children are observed ‘reading’ a picture book to an examiner in a clinic room during the administration of the ADOS, Module 3. This picture book called, Tuesday (Wiesner, 1991) is full of elaborate illustrations and few words. The examiner is ‘scoring’ their narratives from 0 to 2 on eight different imagery-related concepts.

This story-book task has several advantages as a narrative observation sample. First, this assessment was standardized. All of the children were seated in the same clinic
room, reading the same book, by examiners who are trained to provide minimal prompting and direction. Secondly, they are creating a story from a picture-book that provides few words. In this way, the children were required to develop their own narratives, rather than simply reading the story. Finally, the book includes characters with intense facial expressions, and uses vivid illustrations. Thus, the story allows the participants to construct elaborate stories that are full of emotional valence.

There are two significant disadvantages of using the story-book task on the ADOS for the WIOS validation study. The story-book used on the ADOS, *Tuesday*, is whimsical; and follows an unrealistic, inconsistent plot-line. This somewhat odd book about flying frogs may not be accessible or interesting to some of the children. Further, the plot-line does not follow a logical sequence, which may have hindered some of the student’s performance.

Thus, although, the story-telling task was used for the pilot testing and validation tests of the WIOS, the test was designed to be *used on any narrative*. In fact, the test may be more effective if a different story was chosen. Research indicates that a story like the *Three Billy Goats Gruff* would elicit feeling, a sense of character motivation, and a sequence of events (Strickland et al., 1985). However limiting this one story choice may have been, this use of one story has been used effectively in other imagery assessment measures (Sadoski, 1983: used a 5th grade basal reader).

Further necessitating the creation of an observation instrument, this test was standardized on a set of archived video tapes, such that the children could not be asked direct questions about their imagery. Even with this is significantly limiting criterion, the
researcher decided it was still important to prepare an assessment that could be conducted on archived videotapes because these data are routinely collected during any diagnostic assessment. Thus, an assessment of this nature would offer additional data regarding the functioning of the children during routine tests without requiring further assessment. Other studies have used video-taped ADOS administrations to assess language differences in children with autism (Tager-Flusberg, 2006).

*Operationalize the construct to be measured and define test domains.* Now that the purpose of the test has been defined, the Benson and Clark (1982) model provided that test developers must operationally define the construct. This construct, then can be used to define the test domains and to create the item pool. A strong body of literature suggests that children with better imagery produce more coherent, imagery-rich narratives (Bell, 1991b; Cain, 2009; Cain, Oakhill, & Bryant, 2004; Joffe et al., 2007; Norbury, 2002). Essentially, better imagers are better story-tellers. Thus, this instrument was built on the foundation that children’s imagery abilities could be discerned from systemized observation of their story telling.

However, the question of how to measure these abilities is complex. Which skills should be assessed and how will the quality of the narratives be judged or rated? In order to determine this, the literature was reviewed with regard to the domains of integrative-comprehension imagery. An extensive literature search was conducted; and the results are modeled in Figure 3: Theoretical Model of Integrative-Comprehension Imagery. While it was not possible to evaluate every potential domain, the aim of the study was to address
as many essential areas as was feasible based on the data-set of archived video-tapes.

Thus, this list was summarized into six key imagery areas.

Table 2

*Imagery Domains Defined in the Literature*

<table>
<thead>
<tr>
<th>Skill area</th>
<th>Symptom of poor imagery</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Idea</strong></td>
<td>Difficulty with part to whole</td>
<td>Oakhill &amp; Patel, 1991</td>
</tr>
<tr>
<td></td>
<td>Unable to get the ‘gist’ of stories</td>
<td></td>
</tr>
<tr>
<td><strong>Integration</strong></td>
<td>Challenges with putting ideas together across story events</td>
<td>Cain, 2009; Gambrell &amp; Bales, 1986</td>
</tr>
<tr>
<td><strong>Details</strong></td>
<td>Decreased memory for story details</td>
<td>Pressley, 1977</td>
</tr>
<tr>
<td></td>
<td>Following detailed directions</td>
<td>Paivio 1990</td>
</tr>
<tr>
<td></td>
<td>Pairing detail words with images</td>
<td></td>
</tr>
<tr>
<td><strong>Sequencing</strong></td>
<td>Difficulty knowing what came first, next, and last in a story</td>
<td>Clark &amp; Paivio, 1987</td>
</tr>
<tr>
<td><strong>Inferencing</strong></td>
<td>Trouble with ‘gap filling’</td>
<td>Cain et al., 2004</td>
</tr>
<tr>
<td><strong>Coherence</strong></td>
<td>Children with poor imagery fail to form coherent &amp; cohesive representation of story meaning</td>
<td>Kintsch, 1994</td>
</tr>
<tr>
<td></td>
<td>Bell, 1991b</td>
<td></td>
</tr>
<tr>
<td><strong>Connections</strong></td>
<td>Difficulty understanding how one story event impacts another</td>
<td>Cain et al., 2004</td>
</tr>
<tr>
<td></td>
<td>Pearson, in press</td>
<td></td>
</tr>
</tbody>
</table>

In light of the above literature, it was determined that the imagery related skill areas to be assessed were: Main Idea, Integration, and Details. Although Connections was not retained as a unique domain, the connections items were collapsed into the Integration domain. The importance of the Main Idea domain was identified by researchers in imagery (Norbury & Bishop, 2002; Sadoski, 1983) and researchers in autism (Nuske, 2010). The Integration domain was also supported in imagery research (Sadoski, 1983) and in autism research (Diehl et al., 2006). The Details domain was supported by researchers in imagery (Sadoski, 1983) and researchers in autism (Diehl et
al., 2006; Nuske, 2010; Rutherford, Young, Hepburn & Rogers, 2007). The coherence construct (Cain, 2009; Pearson, in press) was captured under the qualitative ‘Global Coherence Rating.’

**Expert Panel and Peer Review Considerations**

Throughout the instrument development process, peer review and expert panels were systematically consulted. The researcher employed a rigorous peer review process in line with best practice in instrument validation (Benson & Clark, 1982; DeVellis, 2003); as well as, in line with the dissertation development process outlined in previous research (Willard & Leffingwell, 2010). The instrument development expert who was regularly consulted is the university instructor for the psychometric theory course. A statistics expert and university professor was consulted with regard to methodology and statistical techniques. An expert Speech/Language pathologist for children with autism was consulted during instrument helped define the test domains. An expert in autism was consulted with regard to the current theories of cognition in autism. Finally, an expert in comprehension strategies was consulted during each phase of instrument development.

After this panel review, it was determined that the six key areas identified in the literature, and feasibly addressed through observation of the videos, would be assessed. Main Idea, Integration, and Details would be the domains of the test. The other areas mentioned were included as items.

*Review literature to determine that test is needed.* At this point in the test development phase, it is important to review all of the relevant literature to ensure that this test does not already exist (Benson & Clark, 1982). A fairly exhaustive list of
imagery tests was included in the literature review chapter. However, a short list of imagery measures includes: the Breakfast Table Questionnaire (Galton, 1880), the Questionnaire Upon Mental Imagery (QMI: Betts, 1909); the Vividness of Visual Imagery Questionnaire (VVIQ: Marks, 1973); Visualizer-Verbalizer Behavior Observation Scale (VVBOS: Leutner & Plass, 1998) Imagery Exercise Questionnaire (IEQ: Epstien, 1980) and The Movement Imagery Questionnaire (MIQ: Hall, 1985). The majority of studies that include an imagery assessment rely on some form of the QMI (Cook, Melamed, Cuthbert, McNeil & Lang, 1988; Danaher & Thoresen, 1971; Hall, Pongrac, & Buckholz, 1985; Hiscock, 1978; Isaac & Marks, 1994; Orr, Pitman, Lasko, & Herz, 1993).

Each of these instruments was considered carefully and it was determined that none were valid for assessing the form of imagery identified in the purpose statement for this instrument. These assessments are outdated, intended for adults, and never tested in a population of children with autism. A final limitation of these instruments for this research is that these are not designed to assess imagery related integrative-comprehension skills. Thus, even though imagery measurements exist, it was determined that the Willard Imagery Observation Scales was necessary as the instrument addresses skills not previously assessed in prior research.

**Identify Target Groups.** Regarding the target groups, the test was designed to measure imagery in school-aged children. The instrument was also developed to be sensitive to the differences between children with autism and other diagnostic categories. As such, was designed to be sufficiently challenging so as to capture the cognitive
differences, but not so difficult that children with autism and a slightly lower IQ would not be able to participate. Thus, the test is deemed appropriate for children with and without autism with an IQ \( \geq 80 \).

**Phase II: Test Construction**

**Develop initial item pool**

During phase I, it was hypothesized that the items would load on three potential latent factors, which might underlie imagery in autism: Main Idea, Integration, and Details. Items 1-4 (what, size, number, where) theoretically would load on the Main Idea domain. Items 5-11 (color, shape, gestures, movement, sound, smell, and taste) would load on the Details domain. Finally, items 12-15 (background, perspective, mood, and time) will load on the Integration domain, based on the theoretical model. The items were derived from a combination of ideas from the literature and the expert panel. Some of the initial items were derived from Bell’s (1991) imagery intervention research. Other items were added to the model in consultation with experts.

**Qualitative Evaluation by Judges**

This initial list of items was reviewed in three ways. First, the researcher brought a small group of videos to an expert in comprehension and asked her to view the videos and score the 15 items from 0 to 2. The expert reviewer was blind to the scores assessed...
by the researcher. The score sheets of each of the judges were compared and discrepancies were discussed. Each item was considered as to whether or not the item was relevant, accurate, and clear (RAC score). This same analysis was conducted with a reliability coder. Again, the reliability coder was blind to the ratings of the researcher; any discrepancies were resolved, and items were revised, as needed, to increase rater agreement. A precise explanation of these revisions is described in the following section and a table of revisions is provided in Table 5.

**Define Scoring Criteria**

Generally, scoring criteria were that 0 point responses indicate that the child did not make any reference to the item. For example, if the participant did not mention any characters, he or she would receive a 0 on the *characters* item. A 1 point response indicated that the child made a vague or unsubstantiated reference to the item or construct. For example, if the participant used the word ‘because’ in attempt to explain a causal relationship but failed to clearly articulate it, he would receive 1 point for *causal connections*. Finally, if the participant clearly demonstrated the skill assessed by the item, he would receive 2 points. For example, if the child said, “He went outside and then flew around for a while, and then went back in the house,” he would receive 2 points for *sequencing*. A full description of the scoring criteria is included in the Appendix.
Phase III: Reliability

During Phase III the instrument was pilot tested for reliability. Pilot testing, overall, included 3 iterations (N=25). The pilot sample was randomly selected from the larger sample of 71 participants. The method of random selection used was stratified sampling; which is known to be as accurate, or more accurate, than simple random sampling (Cochran, 1946). This technique generally reduces sampling error in highly heterogeneous populations (Cochran, 1946). The groups were divided into strata based on IQ (Lower IQ=80-99; Higher IQ=100-140) and diagnosis (Autism, Asperger’s, Typical). Once groups were divided into strata, a simple random sample was taken for the Pilot test (N=25). The remaining participants were included in the validation sample (N=46), for a combined total of the original sample (N=71) (See Figure 6).
Due to the stratified sampling procedure described, the pilot sample was similar to the total sample in most respects. The Mean IQ score was roughly equivalent in the pilot group as compared to the whole sample. The age in years was similar across samples. The composition of girls as compared to boys was roughly equivalent. The diagnostic groupings were similar across samples as well. Thus, the pilot test results should closely resemble that of the entire sample (See Table 3).
Table 3

Pilot Sample Description

<table>
<thead>
<tr>
<th>Description</th>
<th>Entire Sample</th>
<th>Pilot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>IQ</td>
<td>106 (17)</td>
<td>103 (16)</td>
</tr>
<tr>
<td>Age in years</td>
<td>8y (2 y)</td>
<td>9 y (2 y)</td>
</tr>
<tr>
<td>Gender (Boys)</td>
<td>52, 73%</td>
<td>18, 72%</td>
</tr>
<tr>
<td>Autism</td>
<td>28, 40%</td>
<td>10, 40%</td>
</tr>
<tr>
<td>Spectrum</td>
<td>20, 28%</td>
<td>9, 36%</td>
</tr>
<tr>
<td>Typical</td>
<td>23, 32%</td>
<td>6, 24%</td>
</tr>
<tr>
<td>N</td>
<td>71</td>
<td>25</td>
</tr>
</tbody>
</table>

Pilot testing process. The first iteration was an initial reliability test of the items (Version 1). Revisions were made to the items based on the results of these reliability tests, and the scale was tested again (Version 2). During the second iteration, exploratory factor analysis was conducted to assess the factor loadings of each item and the empirical factor structure of the domains. Lastly, the revised scale was tested for reliability and found to be acceptable (Version 3).

For inter-rater reliability, a reliability coder was employed. The reliability coder was trained in autism research, ADOS administration, human subjects ethics, and was approved by the IRB for this study. The reliability coder was trained in three phases for this study. First, the reliability coder watched the researcher code a subset of cases. The reliability coder practiced coding the videos and reliability was calculated (Version 1). The scale was revised due to variability and reliability concerns during pilot testing. The
rater was then re-trained on the new scale and any disagreements were resolved. Finally, inter-rater reliability was re-assessed on the final version of the scale (Version 3).

*Intra-Class Correlations.* The statistical method used for reliability was *intraclass correlation coefficients* (ICC; Pearson, 1901) (McGraw & Wong, 1996; Von Eye & Mun 2005). This method provides researchers with a ratio of the variance between raters compared to the total variance (Lai & Waltman, 2008). Raters are assumed to use a common metric and cases are assumed to be randomly selected (Von Eye et al., 2005). Intra-class correlations are more powerful than straight correlation because the researcher can require absolute agreement (rather than correlation which only requires a consistent pattern). The assumptions of random selection and common metric were met. Inter-rater reliability was calculated on 25% of the total cases and found to be strong.

*Reliability criteria.* The acceptable standard for reliability is unsettled in the psychometrics research (Carmines & Zeller, 1979). The only clear standard is that the reliability statistic must be reported (Carmines & Zeller, 1979). Generally, a standard of .70 is used, which is based on the standard set by Nunnally, expert in reliability (Bobko, 2001, Nunnally, 1994). However, researchers in the social sciences have reported that the requirement for more subjective measures is .60; and the standard for high-stakes testing with objective measures is .80 (Kubiszyn, & Borich, 2010). In this study, the reliability of the factors was required to be higher than .60; and the standard for inter-rater reliability was .80.

*r-i-t’s Reliability:* Another test for internal consistency is to consider the relationship between each item and the total score. The theory is that if all of the items
are measuring some common construct or trait; then each item \( (i) \) is expected to correlate with the total score \((T)\) (Bobko, 2001). This form of reliability is thus called ‘r-i-t’s’ \( (r_{iT}) \). Bobko (2001) reports that, if all items that do not correlate with the total are deleted; internal consistency is ‘guaranteed’ (p.72). This \( r_{iT} \) analysis was conducted on Version 3, and all items were statistically significantly correlated with the WIOS total \((p<.05)\) Thus internal-consistency reliability was demonstrated. The results of the other reliability tests are as follows in Table 4.
Table 4

Scale Reliability

<table>
<thead>
<tr>
<th></th>
<th>Version 1</th>
<th>Version 2</th>
<th>Version 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-Rater Reliability (ICC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard used = &gt;.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entire Scale</td>
<td>.73</td>
<td>*</td>
<td>.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main Idea Domain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration Domain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Details Domain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal-Consistency (α)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard used = &gt;.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entire Scale</td>
<td>.78</td>
<td>.71</td>
<td>.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main Idea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.42</td>
<td>.70</td>
</tr>
<tr>
<td>Integration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.68</td>
<td>.61</td>
</tr>
<tr>
<td>Details</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.18</td>
<td>**</td>
</tr>
<tr>
<td>Problematic Items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (what) 5 (color), 6 (shape), 12 (background), &amp; 13 (perspective)</td>
<td>2 (emotions), 9 (integrative statement), 10 (setting), 11 (size), 12 (time of day), 14 (positioning)</td>
<td>None.</td>
<td></td>
</tr>
</tbody>
</table>

- In version 1, it was not possible to calculate domain reliability because of limited variance.

* In version 2, inter-rater reliability was not calculated because the scale was in the process of revision.
Reliability results.

Version 1. Results indicate that the judges’ ratings were reliable (ICC=.73), and the internal consistency was fair (α=.78). Reliability of the domains could not be calculated because of limited variability in item scores. Items: 1, 5, 6, 12 and 13 appeared to be problematic because reliability would increase if the items were deleted. Both a statistical and substantive evaluation of these items ensued for these items (see Table 5).

Version 2. The results of Version 2 indicate that reliability of the revised scale was adequate (α=.71). The Main Idea domain showed inadequate reliability (α=.42); as did the Integration domain (α=.68), and the Details domain (α=.18). There were several problematic items within the Details domain. Items were assessed to be problematic if reliability would increase if the item was deleted. The Details domain and all of the items within it were deleted (see domain revisions). Within Benson & Clark’s (1982) model, it was necessary to revise any unreliable items and then retest the instrument. In consultation with a subject matter expert, the items were heavily revised (see Table 5).

Version 3. Version 3 showed adequate reliability for all of the factors and items. Inter-rater reliability of the entire scale was strong (ICC=.93). Inter-Rater Reliability of the domains was good: Main Idea domain (ICC=.94) and Integration Domain (ICC=.86). Internal-consistency reliability of the scale was fair (α=.72). The domains showed adequate reliability: Main Idea (α=.70) and Integration (α=.61). There were no problematic items. Thus, all of the items were retained and Version 3 became the final scale.
Table 5

*Item Revisions to Enhance Reliability*

<table>
<thead>
<tr>
<th>Version 1</th>
<th>Version 2</th>
<th>Version 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. What</strong>&lt;br&gt;Gives an explanation of who or what the story is about.</td>
<td>Item replaced with two items:&lt;br&gt;1) Characters&lt;br&gt;4) Actions</td>
<td>1) Characters item was retained under Main Idea</td>
</tr>
<tr>
<td><strong>2. Size</strong>&lt;br&gt;Gives a sense of dimension or size of objects in the story</td>
<td>Scoring criteria was loosened. Item collapsed into:&lt;br&gt;11) Size or shape.</td>
<td>Size item was deleted due to insufficient reliability.</td>
</tr>
<tr>
<td><strong>3. Number</strong>&lt;br&gt;Tells how many objects or characters there are</td>
<td>Retained but scoring criteria was loosened. Moved to Details domain:&lt;br&gt;13) Number</td>
<td>Number item was deleted due to insufficient reliability.</td>
</tr>
<tr>
<td><strong>4. Where</strong>&lt;br&gt;Gives a sense of place or setting</td>
<td>Item was changed to:&lt;br&gt;12) Setting</td>
<td>Where/Setting was deleted due to insufficient reliability</td>
</tr>
<tr>
<td><strong>5. Color</strong>&lt;br&gt;Tells about how objects look or uses color words</td>
<td>Item was collapsed into:&lt;br&gt;9) Sensory (see explanation for adding sensory in the explanation that follows)</td>
<td>Color/Sensory was deleted due to insufficient reliability and item variance</td>
</tr>
<tr>
<td><strong>6. Shape</strong>&lt;br&gt;Gives a sense of the shape of any object in the story</td>
<td>Item was refined and collapsed into:&lt;br&gt;11) Size or Shape</td>
<td>Shape/Sensory was deleted due to lack of insufficient reliability and variance</td>
</tr>
<tr>
<td><strong>7. Gestures</strong>&lt;br&gt;Gestures must be a clear attempt to communicate.</td>
<td>Scoring criteria was tightened. Item moved to Integration domain:&lt;br&gt;6) Gestures</td>
<td>Gestures was retained under the Integration Domain</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Version 1</th>
<th>Version 2</th>
<th>Version 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.</td>
<td><strong>Movement</strong> Describes any character moving, in motion, or performing an action</td>
<td>Item was replaced with: 4) Actions</td>
<td>Item was retained but moved to Main Idea domain.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2) Actions</td>
</tr>
<tr>
<td>9.</td>
<td><strong>Sound</strong> Describes any noise with the ears. Does not have to be specific</td>
<td>Item was collapsed into: 9) Sensory (see explanation that follows)</td>
<td>Item deleted due to insufficient variance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td><strong>Smell</strong> Any reference to how something smelled or to the nose.</td>
<td>Item was collapsed into: 9) Sensory</td>
<td>Item deleted due to insufficient variance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td><strong>Taste</strong> Any reference to the taste of food. Does not have to be specific</td>
<td>Item was collapsed into: 9) Sensory</td>
<td>Item deleted due to insufficient variance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td><strong>Background</strong> Gives an overall sense of place and time. Sets the stage for the story.</td>
<td>Item was collapsed into: 10) Setting or Background</td>
<td>Item deleted and new item created.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8) Integrative Statements</td>
</tr>
<tr>
<td>13.</td>
<td><strong>Perspective</strong> Gives a sense that the storyteller understands the character’s point of view.</td>
<td>Scoring criteria was tightened. Another item was added: 13) Perspective Taking 14) Positioning/Placement</td>
<td>Item retained under Integration Domain: 7) Perspective Taking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*Positioning / Placement was deleted.</td>
</tr>
<tr>
<td>14.</td>
<td><strong>Mood</strong> Gives overall emotional valence, or uses feeling words</td>
<td>Item was changed to: 1) Emotions</td>
<td>Item retained under Integration Domain: 4) Emotions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td><strong>When:</strong> Clock time or temporal order</td>
<td>Item was collapsed into: 3) Sequence 12) Time of Day</td>
<td>Sequence item retained under Main Idea Domain: 3) Sequence</td>
</tr>
</tbody>
</table>
Item revisions based on reliability tests. As shown in Table 5, there were a variety of changes made to the items, in order to increase the reliability of the scale. A manual of the new items is provided in Appendix A. This section is a brief summary of the changes made to Version 3 items; with an associated rationale for each change.

Item 1: characters. In version 1, item 1 was called ‘what’ and was intended to measure the main events and characters of the story. Early pilot testing revealed that the item showed limited variance. This item was then revised and collapsed into two items: characters and actions. The characters item shows strong support in the imagery literature (Sadoski, 1983) and the autism literature (Rutherford, et al., 2007). The scoring criteria was tightened (increased in difficulty) such that the examinee was required to mention 2 animal characters for a 1 point; or 2 animals and a human character to receive a 2 point score. The characters item showed strong reliability during pilot testing and was thus retained in the final scale: Version 3.

Item 2: actions. The actions item has support in the literature (Rutherford, et al., 2007; Strickland & Feeky, 1985). As stated, this construct was originally captured under the ‘what’ item; intended to capture the major events of the story. However, qualitative evaluation by judges revealed that the item was difficult to score because it lacked specificity. Thus what was changed to actions in Version 2. However, pilot testing of Version 2 revealed that the item showed insufficient reliability. In order to address this, scoring criteria were tightened, and the clarity was enhanced. The final item, ‘actions’ in Version 3 of the scale was an assessment of the child’s ability to name at least two actions (1 point). A 2 point response required the child to describe the characters as actors.
on the environment (See Theoretical Model, Figure 3). Reliability tests (using the r-i-t’s method) showed that adequate reliability. Thus, the item was retained in the final scale: Version 3.

**Item 3: sequence.** The sequence item was initially captured under ‘time of day’ which was intended to be a measure of temporal order. The ‘time of day’ item showed poor reliability (both inter-rater and internal consistency reliability). A qualitative evaluation in consultation with experts revealed that sequence is more important to imagery than ‘time of day.’ Imagery research supports the validity of the sequence item (Bell, 1991b; Clark & Paivio, 1987). There is also support in the literature for children with autism showing poor sequencing skills (Rutherford, Young, Hepburn, & Rogers, 2007). The sequence item scoring was defined such that one sequential connection would receive 1 point and two sequential connections would receive 2 points. The sequence item showed strong reliability during pilot testing and was retained in Version 3.

**Item 4: emotions:** The emotions item was initially captured under ‘mood.’ However, the item showed poor inter-rater reliability because it was unclear whether the item was referring to the overall spirit and mood in the story or to specific emotions of the characters. A child might have mentioned the overall mood of the story by saying something like ‘silly’ or ‘this is a weird story.’ Another child might provide an emotion word such as ‘happy.’ Still another child might have given a more specific explanation of mood by saying, “He feels sad because he can’t fly with his friends.” In light of this, it was determined that the mood item would be re-named emotions and would be scored differently. A 1 point response required the participant to use 2 feeling words such as
‘happy’ or ‘scared.’ A 2 point response required the child to provide a rationale for that feeling. This item showed limited reliability in Version 2 but when scoring criteria were tightened in Version 3, the item demonstrated acceptable reliability.

**Item 5: gestures:** A careful review of the literature ensued to determine whether or not gesture use should be included. Literature revealed that gestures are considered important to the study of imagery (Bell & Bonetti, 2006). Bell and colleagues state that, “A good storyteller links imagery with language through gestures (p.27)” Further, gesture use has been shown as a pragmatic language impairment in the autism population in prior research (Philofsky, Fidler & Hepburn, 2007). Children with autism have difficulties understanding the gestures of others and using gestures appropriately in their own communication. As such, it was decided that gestures should be retained in the scale and would stand alone as a unique item.

**Item 6: cause and effect:** Cause and effect was added in version 2. This item was added to the Integration domain due to literature indicating that sequential and causal connections are important to comprehension (Pearson, in press; Strickland & Feeky, 1985) and imagery (Cain, Oakhill, & Bryant, 2004). The reliability coder was trained on this item and scoring criteria were clarified in the manual. Scoring required that a child make a clear causal connection to receive 2 points, and a vague reference to causal connections to receive 1 point. The item showed acceptable internal-consistency reliability and inter-rater reliability and thus was retained in Version 3.

**Item 7: perspective taking.** This item was intended to measure the child’s ability to assess where objects were in the story in relation to other items. For example, the child
might say that the tree was next to the house or that the birds flew over the chimney. The problem was that it was unclear as to whether the item should be measured from the child’s perspective or the character’s perspective. Further, the item was not clear as to whether it was measuring the character’s physical perspective or mental state.

A review of the literature reveals that taking on the mental state of characters in a story is important to imagery assessment (Sadoski, 1983) and to assessment of integrative comprehension in autism (Nuske, 2010; Rutherford et al., 2007). It was determined that this item would be re-named item 7: perspective taking. Revised scoring criteria were that a response that included the physical perspective of the character would receive 1 point. A response that included the character’s mental state would receive 2 points. This item showed adequate reliability and was thus retained in Version 3.

**Item 8: integrative statements.** A visual inspection of the child’s responses in Version 1, revealed that some children provided an overall statement that summed up the story or approached a deeper level of synthesis. For example, one child said, “That’s neat that they can fly. I wish I could fly.” As such, item 8: integrative statements was added to Version 2 of the scale. Participants received 1 point for any statement made during the story and 2 points for making a statement at the end that served to pull the story together. The item showed insufficient reliability in Version 2 but revisions were made and it became reliable in Version 3 so it was retained.

**Sensory.** A problem of insufficient variability in item scores occurred on several items within the sensory category. In Version 1, all of the participants scored a 0 on the sensory items (color, shape, sound, smell, and taste). It was determined that there was not
sufficient information about sensory details like sounds smells and tastes, provided in the particular story. In light of this problem with variability, all of these items collapsed into: Item 9: sensory (Version 2). The logic for including a sensory item was that sensory awareness is known in the literature to bear a strong relationship with imagery (Bell & Bonetti, 2006). Although this item demonstrated strong reliability, it was the only item that loaded onto the Details domain. Thus, when the Details domain was deleted, this item was removed from the scale. This requirement of deleting sensory from the scale is unfortunate because of the strong research base for this construct, and invites opportunity for future research in this area (See Implications for Research).

Global Coherence. During pilot testing, it was revealed that the scale had previously failed to capture the overall coherence of the narrative. As shown in Figure 3, the theoretical model included a minor branch called coherence or cohesive representation. A child with poor imagery might tell stories that lack a consistent plot-line (Bell, 1991b; Cain, Bryant & Oakhill, 2004; Clark & Paivio, 1987; Kintsch, 1988; Kinsch, 1994), a description of the characters (Sadoski, 1983), or a climax (Sadoski, 1983). Pilot testing revealed that some of the participants, particularly those with Asperger’s, scored fairly well on the WIOS because they had covered all of the areas being measured; however, the narrative was still lacking in overall coherence.

In order to account for this, a simple qualitative rating was added in Version 3 of the scale. This Global Coherence rating was scored from 0-100. This rating was tested for reliability using the r-i-t’s method and found to be strong (r=.83**). However, inter-rater reliability and Cronbach’s alpha was not calculated because Global Coherence was not
part of the WIOS item set; but rather, a separate scale. This rating was different because it was based on the examiner’s overall impressions of the narrative, rather than the narrow skills assessed by items. Future researchers might consider expanding the Global Coherence scale; including new items to capture the construct, and forming a unique domain (See Implications for Research).

**Factor Analysis Results.** The goal of factor analysis was first to first understand how many latent factors of imagery underlie the items. Once the factors were defined, the factor loadings of each item could be assessed. The criterion used was >.35, which is a conservative estimate for item factor loadings during scale development (Sadoski, 1983).

*Version 1.* Version 1 of the test was based on the theoretical factor structure identified in the literature (See Figure 3). This factor structure included 3 theoretical domains (Main Idea, Integration, and Details). The domains could not be tested on Version 1, because most items showed inadequate variance. Instead, the items were revised to enhance reliability and variance; and then retested in Version 2.

*Version 2.* In order to decide on the number of factors to be used, an Exploratory Factor Analysis (EFA) was conducted. The EFA identified 6 potential factors. Results indicate that a three factor model would explain 49% of the variance; a four factor model would explain 60%; five factors explain 70%, and six factors explain 78%. The statistical 6 factor solution presented a problem. Some of the items (within the Details domain) were loading on factors 4, 5, & 6 which were not included in the model. Further empirical tests were conducted on the Details domain and it was deleted in Version 3 (see next section for the rationale).
Version 3. Once all of the item revisions had been completed and the 2 factor model had been established, eigenvalues were calculated to determine the percent of variance explained by the model. Next, factor loadings were tested (see Table 7). Version 3 was assessed to show adequate reliability and an empirically valid factor structure. The factor analysis results of Version 3 are provided in the next section.

Domain Revisions. The table that follows provides a summary of the changes that were made to the domains based on the empirical factor structure. Each item was assessed during the EFA, both empirically and substantively. If the factor loading was consistent with theory (as in item 1: characters), it was retained under that factor in Version 3. If the item loaded empirically on another factor in the model (as in item 3: sequence), a qualitative evaluation ensued. The item was moved if there was a theoretical justification. For example, the sequence item was originally thought to fall under Integration; however, the EFA revealed that it loaded on the Main Idea domain. A qualitative evaluation revealed that indeed sequencing events should be included in the Main Idea domain. Lastly, if the item loaded on an unidentified domain, it was deleted.

Factor Analysis Results. The goal of factor analysis was first to first understand how many latent factors of imagery underlie the items. Once the factors were defined, the factor loadings of each item could be assessed. The criterion used was >.35, which is a conservative estimate for item factor loadings during scale development (Sadoski, 1983).

Version 1. Version 1 of the test was based on the theoretical factor structure identified in the literature (See Figure 3). This factor structure included 3 theoretical domains (Main Idea, Integration, and Details). The domains could not be tested on
Version 1, because most items showed inadequate variance. Instead, the items were revised to enhance reliability and variance; and then retested in Version 2.

**Version 2.** In order to decide on the number of factors to be used, an Exploratory Factor Analysis (EFA) was conducted. The EFA identified 6 potential factors. Results indicate that a three factor model would explain 49% of the variance; a four factor model would explain 60%; five factors explain 70%, and six factors explain 78%. The statistical 6 factor solution presented a problem. Some of the items (within the Details domain) were loading on factors 4, 5, & 6 which were not included in the model. Further empirical tests were conducted on the Details domain and it was deleted in Version 3 (see next section for the rationale).

**Version 3.** Once all of the item revisions had been completed and the 2 factor model had been established, eigenvalues were calculated to determine the percent of variance explained by the model. Next, factor loadings were tested (see Table 7). Version 3 was assessed to show adequate reliability and an empirically valid factor structure. The factor analysis results of Version 3 are provided in the next section.

**Domain Revisions.** The table that follows provides a summary of the changes that were made to the domains based on the empirical factor structure. Each item was assessed during the EFA, both empirically and substantively. If the factor loading was consistent with theory (as in item 1: *characters*), it was retained under that factor in Version 3. If the item loaded empirically on another factor in the model (as in item 3: *sequence*), a qualitative evaluation ensued. The item was moved if there was a theoretical justification. For example, the sequence item was originally thought to fall under
Integration; however, the EFA revealed that it loaded on the Main Idea domain. A qualitative evaluation revealed that indeed sequencing events should be included in the Main Idea domain. Lastly, if the item loaded on an unidentified domain, it was deleted.

Table 6

*Domain Changes Based on Factor Structure*

<table>
<thead>
<tr>
<th>Version 1</th>
<th>Version 2</th>
<th>Version 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Idea</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What</td>
<td>Characters</td>
<td>Characters</td>
</tr>
<tr>
<td>Size</td>
<td>Actions</td>
<td>Actions</td>
</tr>
<tr>
<td>Number</td>
<td>Gestures</td>
<td>Gestures</td>
</tr>
<tr>
<td>Where</td>
<td>Emotions</td>
<td>Emotions</td>
</tr>
<tr>
<td><strong>Integration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background</td>
<td>Sequence</td>
<td>Emotions</td>
</tr>
<tr>
<td>Mood</td>
<td>Causal Connections</td>
<td>Causal Connections</td>
</tr>
<tr>
<td>Perspective</td>
<td>Perspective Taking</td>
<td>Perspective Taking</td>
</tr>
<tr>
<td>Time</td>
<td>Integrative Comment</td>
<td>Integrative Statements</td>
</tr>
<tr>
<td><strong>Details</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td>Sensory</td>
<td>*</td>
</tr>
<tr>
<td>Shape</td>
<td>Size or shape</td>
<td>*</td>
</tr>
<tr>
<td>Gestures</td>
<td>Setting</td>
<td>*</td>
</tr>
<tr>
<td>Movement</td>
<td>Time of day</td>
<td>*</td>
</tr>
<tr>
<td>Sound</td>
<td>Number</td>
<td>*</td>
</tr>
<tr>
<td>Smell</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taste</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Domain revisions: Details domain.* As shown on the previous table (Table 6), the Details domain was deleted in Version 3. This decision was made based on the results of empirical and qualitative analysis. During reliability testing, the domain showed poor internal consistency (Chronbach’s alpha=.18). During the EFA, it was determined that the detail items were not loading onto the Detail domains. The only item that loaded on the Details domain was item 9 *sensory* (factor loading=.91). However, item 10 *setting*; item
11 *size*; item 13 *number* and item 14 *positioning/placement* all loaded onto factors 4, 5, or 6, which were not included in the model. Thus, the empirical factor structure did not support the Details domain.

*Qualitative Analysis.* Before deleting the domain, a qualitative evaluation was conducted in consultation with subject matter experts. The intent of the domain was to show whether or not the children with autism provided irrelevant, somewhat random details, consistent with past research (Nuske, 2010; Wahlberg, 2001). However, factor analysis techniques split out factors based on the clustering of similar items. In this case, the items were all intentionally not related because the construct was intended to measure miscellaneous details. Based on the results of empirical and qualitative analysis, the domain was deleted. This was unfortunate because the domain has support in the literature (Nuske, 2010; Norbury & Bishop, 2002). There may be a way to better capture Details in future research (see Implications for Research section).

*Principal Components Analysis.*

The purpose of this phase of domain evaluation was to determine the empirical basis for the two factor model identified during Exploratory Factor analysis. If indeed Version 3 held up as demonstrating an empirically valid factor structure, Version 3 could be retained as the final version and pilot testing would be complete. If, however, the factor loadings were again inconsistent with theory, the items would need to be revised again, or new factors would need to be identified.

This factor analysis was conducted using the Principal Components Analysis, varimax rotated solution (Gable, 1993). The rotated solution is virtually always used in
psychometrics for test development (C. Nelson, *personal communications*, April 15, 2012). The varimax rotation is used to ‘split out’ the factors into unique constructs based on the clustering of items (Gable, 1993). The varimax rotation is used when factors are assumed to be uncorrelated because it is an orthogonal solution (Gable, 1993). The results of this analysis were as follows in Table 7.

Table 7

*Factor Loadings for Version 3*

<table>
<thead>
<tr>
<th></th>
<th>Integration</th>
<th>Main Idea</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Characters</td>
<td>.10</td>
<td>.88</td>
</tr>
<tr>
<td>2. Emotions</td>
<td>.78</td>
<td>-.083</td>
</tr>
<tr>
<td>3. Gestures</td>
<td>.53*</td>
<td>.56*</td>
</tr>
<tr>
<td>4. Actions</td>
<td>-.32</td>
<td>.58</td>
</tr>
<tr>
<td>5. Sequence</td>
<td>.30*</td>
<td>.80*</td>
</tr>
<tr>
<td>6. Causal connections</td>
<td>.64</td>
<td>.28</td>
</tr>
<tr>
<td>7. Perspective Taking</td>
<td>.54</td>
<td>.04</td>
</tr>
<tr>
<td>8. Integrative Statements</td>
<td>.71</td>
<td>.07</td>
</tr>
</tbody>
</table>

*Note. Criteria for factor loading = >.35

*Cross-loadings on both factors. Item 3 was retained in the Integration Domain based on the previous substantive analysis. Item 5 was retained in the Main Idea domain due to a higher statistical factor loading and the theoretical analysis previously described.*
Factor Loadings for Main Idea Domain for pilot study (N=25). Regarding factor loadings, several of the items corresponded to the theoretical factors. The criterion was that any factor loading greater than .30 was decidedly loading on that statistical factor. First, item 1 characters loads on the Main Idea Domain, which is consistent with the theory (.88). Item 4 actions loads on the Main idea Domain (.58). The sequence item loads on the Main Idea domain (.80). Although items 4, actions; and item 5, sequence; were not anticipated, this statistical finding makes sense based on theory (see Domain revisions).

Factor loadings for the Integration Domain for pilot study (N=25). The statistical factor structure revealed that item 2, emotions, loads on the Integration domain (.78), rather than the Main Idea domain. This loading, though unexpected, was sensible based on a substantive evaluation. The statistical factor structure revealed that item 2, emotions loads on the Integration domain, rather than the Main Idea domain. This loading, though unexpected, is sensible in light of the fact that the two point responses require some integration of the facts in the story with the emotion that is expressed (children are required to provide a rationale for why a character feels a certain way based on the story).

Also inconsistent with the expected factor structure, item 3: gestures, cross-loads on Main Idea and Integration domain. Gesture use seems to be captured, to some degree in both domains. Children who understand the Main Idea have created a coherent mental representation of the story; and thus, are more likely to describe that image with gestures. Gestures are generally used as an attempt to help the listener understand the thoughts and images of the speaker. Imagery researchers, Bell & Bonetti explain, “a good story teller
links images with language through gestures” (p.27). However, gesture use also involves a degree of integration because information about the characters must be integrated with the story’s events in order to provide a basis for using a descriptive gesture. Based on the substantive evaluation, gestures was moved to the Integration domain. Future researchers might consider adding gestures to a Coherence domain that would include the global coherence item, *gestures*, plot, and climax items. This might more adequately capture the construct. In the current WIOS model; however, *gestures* seems to fit reasonably well within the Integration domain.

Taken together, the remaining items in the Main Idea domain, based on the factor loadings and substantive evaluations were: characters, actions, and sequence. A careful evaluation of the new Main Idea domain, indicated that the construct is valid. The expert panel evaluation had determined that a child who has provided a narrative that includes the characters, the actions of the characters, and the sequence of events; has clearly demonstrated understanding of the Main Idea.

Consistent with theory, items 6, *causal connections*; 7 *perspective taking*, and 8 *integrative statements* loaded on the theoretical factor structure under the Integration Domain. The factor loadings were: *causal connections* (.64) *perspective taking* (.54) and *integrative statements* (.71). Thus, these three items could remain in the Integration domain based on theory and empirical analysis. Thus, the Integration domain included: *emotions, gestures, causal connections, perspective taking, and integrative statements*.

*Variance explained by 2 factor model in pilot study (N=25).* This new two factor model has an unfortunate consequence, statistically. First, the inclusion of fewer items in
the scale tends to reduce reliability (DeVellis, 2003). Secondly, the two factor model explains less variance overall. The eigenvalues for the two-factor model are provided in Table 8, below. The two factor model explains 55% of the variance in the imagery construct, leaving 45% of the variance unexplained by the model. As previously described, there are more than six factors of imagery that could have potentially been included in the instrument. However, due to measurement constraints, inclusion of these items was disallowed. Future researchers might consider adding items from the Theoretical Model (Figure 3), which would likely explain more variance in imagery skills (see Implications for Research).

Table 8

*Factor Analysis for Two Factor Model*

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28.579</td>
</tr>
<tr>
<td>2</td>
<td>55.358</td>
</tr>
</tbody>
</table>

The final scale after all of the pilot test results in two factors: Main Idea and Integration. There are a total of 8 items on the scale. The scale (Version 3) shows adequate reliability. The final scale items and domains are provided in Table 9. The scoring manual for the final scale is shown in the Appendix.
Table 9

*Final Scale*

<table>
<thead>
<tr>
<th>Main Idea Domain</th>
<th>Integration Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Characters</td>
<td>4. Emotions</td>
</tr>
<tr>
<td>2. Actions</td>
<td>5. Gestures</td>
</tr>
<tr>
<td></td>
<td>7. Perspective Taking</td>
</tr>
<tr>
<td></td>
<td>8. Integrative Statements</td>
</tr>
</tbody>
</table>

**Phase IV: Validity**

- Readminister instrument to validation sample (N=46)
- Analyze psychometric properties of the scale
- Continued validity tests

*Re-administer instrument to validation sample.*

The scale had demonstrated reliability with the pilot sample of 25 participants. However, it was important to re-run this analysis on a different sample in order to assess validity of the instrument. This second sample was randomly selected from the same archival data set. As stated, the validation sample was randomly selected, first using a stratified random sample based on IQ and diagnosis. Next, the participants were selected from the strata using a simple random sampling approach.
Figure 7

**Validation Sample**

![Diagram showing the combination of Pilot sample (N=25) and Validation sample (N=46) to form the Total sample (N=71)]

*Analyze Psychometric Properties of Scale:* Identical procedures were used to test the reliability of the scale on the validation sample. Reliability was calculated for each domain using internal consistency reliability (Chronbach’s alpha). Factor analysis was conducted to determine the variance explained by a two factor model and the factor loadings of the items. Results were as follows in Table 10.
Table 10

Psychometric Properties

<table>
<thead>
<tr>
<th>Validation Sample (N=46)</th>
<th>Internal-Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard used = α &gt; .60</td>
<td></td>
</tr>
<tr>
<td>Entire Scale</td>
<td>.72</td>
</tr>
<tr>
<td>Main Idea</td>
<td>.68</td>
</tr>
<tr>
<td>Integration</td>
<td>.66</td>
</tr>
</tbody>
</table>

Problematic Items (slightly increased alpha if deleted)
2 (actions), 4 (emotions)

<table>
<thead>
<tr>
<th>Factor Loadings</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard used = &gt; .30</td>
<td></td>
</tr>
<tr>
<td>Main Idea</td>
<td></td>
</tr>
<tr>
<td>1.Characters = .89</td>
<td></td>
</tr>
<tr>
<td>2.Actions = .64</td>
<td></td>
</tr>
<tr>
<td>3.Sequence = .75</td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td></td>
</tr>
<tr>
<td>4.Emotions = .69</td>
<td></td>
</tr>
<tr>
<td>5.Gestures = .30</td>
<td></td>
</tr>
<tr>
<td>6.Causal Connections = .74</td>
<td></td>
</tr>
<tr>
<td>7. Perspective Taking = .79</td>
<td></td>
</tr>
<tr>
<td>8. Integrative Statements = .58</td>
<td></td>
</tr>
</tbody>
</table>

Results indicate that the scale shows adequate reliability overall (Cronbach’s alpha = .72). The scale included 8 items and the utilized the same two factor model from Phase IV. Even though item 2 and item 4 are slightly problematic, as reliability would increase if deleted, reliability would only increase to .73 or .74, respectively.

Substantively, these items hold up and contribute to the factors. Further, all of the items had shown sufficient reliability through the r-i-t’s reliability test, and were seen as valid items by the expert panel. Therefore, all 8 items and both domains were retained.
Factor Analysis. The Eigenvalue for the 2 factor solution was virtually the same as in the reliability tests (Eigenvalue=1.6; 55% Variance explained); and the validity sample (Eigenvalue=1.7, 57% Variance explained). Thus, the 2 factor solution explains roughly 57% of the variance in the imagery scale. The factor loadings were stronger in the validation sample than in the reliability sample. With the exception of the gestures item, all of the factor loadings were greater than .50 (criteria = >.30). Thus, the factors seem to be reliable and the factor loadings are consistently loading onto their associated domains.

Continued Validity Tests

Validity Generalization: The scale held-up as consistently reliable in Pilot 3 and in the Validation Sample. It shows reliability of .72, overall. This is a promising result because the pilot tests were run on a different sample (N=25), with the same inclusion criteria as the validation sample (N=46). Both the pilot sample and the validation sample were randomly selected (using stratified random sampling) from the pool of video-tapes in the archival data set. Thus, the scale seems to adequately measure ability across different samples. Thus, there is some evidence that the results of this instrument on the validation samples may be generalizable to similar populations. There was a large population of children with clinically diagnosed autism between the two samples, which is very hard to find and rarely achieved in other studies; thus, strengthening the generalizability of these findings. However, in order to truly test the generalizability of these results, larger samples which are representative of the entire population of kids aged 6-12 in the general population (potentially using census data) would be necessary.
**Face Validity:** Face validity measures whether the test looks like it measures what it was intended to assess (Bobko, 2001). Bobko, statistics expert, explains it thusly, “Thus, face validity is just a subjective judgment that the test accurately reflects the thing you’re trying to measure” (p.74). This form of validity was rigorously tested through the expert panel evaluation which included SME’s from the fields of psychometrics, autism, speech / language pathology in autism, psychoeducational assessment, and comprehension. The panel of experts verified that the items demonstrate adequate face validity.

**Content Validity:** Content validity was also evaluated. Content validity measures the test developer’s ability to write items that reflect a representative sample of the population or domain (Bobko, 2001). Bobko provides that there are 3 criterion used for content validity:

1) how well the universe of interest has been defined, 2) how systematically one has sampled from that universe, and 3) how accurately the items reflect the sampled domains (p.75).

In this study, there were originally three domains that were systematically narrowed down to two, thus satisfying the first condition for content validity. The original scale was comprised of 15 items, summarized into these domains. A systematic review was conducted to assess: the literature on the item, the reliability of that item within the scale, and the scoring criteria. Each item was uniquely assessed on these factors and poor items were deleted, which resulted in the total of 8 items in the final scale. This systematic item sampling meets the second criteria for content validity.
Finally, the third criteria regarding the accuracy of the items in representing the domains, was met through content evaluation by subject matter expert judges. Overall, the test demonstrates content-validity on the 3 aforementioned content validity criteria.

*Messick’s Validity:* The test was evaluated for ‘Messick’s Validity’ which assesses the fairness of the uses of the test (Thorndike, 2005). Messick was a research scientist who was concerned about the social consequences; intended and unintended, of a particular test. For example, a decade ago, students could be placed into special education classes based on the results of one IQ score. This practice has mostly been eradicated in favor of the Response to Intervention Model (Riley, 2011). The problem was that the social consequences of this test were too extreme to be considered valid.

In this study, The Willard Imagery Observation Scales is valid based on Messick’s validity criteria. The instrument is used to provide additional data regarding potential cognitive difficulties experienced in children with autism or typically developing children. These questions were specifically designed to be sensitive to fine differences in the imagery skills of children with autism based on known symptoms. The WIOS, thusly, would give information to clinicians about potential deficit areas that could be targeted and addressed; such as identifying emotions, perspective-taking, making causal-connections, or understanding the sequence of events. Thus, the social consequences of the WIOS are acceptable as the children tested with this instrument are extremely unlikely to be harmed in any way by the results. Rather, the results of this test are expected to help clinicians assesses specific deficits in this population in order to provide support to these children.
**Criterion-related validity:** In order to test criterion validity the results of this test would need to be correlated with the performance criteria it is designed to predict. Indeed, most researchers use the correlation between the test and the predicted performance results to be the test’s validity (Bobko, 2001). It is unknown whether or not the WIOS demonstrates criterion validity. In order to test this, it would be necessary to compare results of this instrument to the results of standardized reading comprehension assessments. In this study, it was not possible to assess the correlation between the WIOS and a standardized comprehension assessment because the data were not available.

However, past imagery researchers have found that the relationship between imagery measures and standardized testing is not significant (Sadoski, 1983). One hypothesis for this finding is that the tests are highly verbally-loaded and thus do not assess the types of skills measured on imagery-related comprehension tests (1983). Thus, although the next step to establish criterion-related validity would be to compare the results of the WIOS to a standardized comprehension instrument, it is possible that it will be difficult to uncover an instrument that could effectively be used for this purpose.

**Construct Validity:** Construct validity was assessed first as a correlation with diagnosis, which will be provided in the next chapter (Chapter 5: Multivariate Analysis). There is some evidence for construct validity in that correlation between diagnosis and scores on the WIOS total is high and significant \((r=0.70; p>.001)\). The domains of this test were consistent with other imagery and autism factor analysis studies (Nuske, 2010; Sadoski, 1983;). However, further tests for construct validity could be demonstrated if the domains of this test were compared to other imagery or comprehension instruments.
Factor Analysis: Preliminary Confirmatory Factor Analysis tests were conducted. In the pilot, the model fit appeared to be somewhat adequate due to a non-significant $\chi^2$ fit index (chi squared is a measure of bad fit so non-significant results indicate good fit). However, the other model fit indexes (CFI, RMR, and RMSEA) did not indicate good model fit. Further, this same test was run on the confirmatory sample. Results were that none of the model fit indexes appeared to be adequate (chi squared, CFI, RMR, or RMSEA). It is unknown why this model fit has not been demonstrated. It is possible that the items are good but some of the factor structure may not be adequate to explain the variance in the latent factor of imagery. It is also possible that there are simply not enough items or factors to provide a best fitting model of imagery. Adding items and factors to the model in order to test this, was not possible on the available archived data set. The researcher acknowledges this limitation of this study and hopes to continue this research in the future (see Limitations section of the Discussion chapter).

Summary of Instrument Validation Results

The Willard Imagery Observation Scales were tested for reliability and validity using Benson & Clark’s (1982) instrument development model. Phase I required that the purpose of the test and the domains of the test be carefully researched and defined. Phase II was the test construction phase. During this analysis, an initial item pool was developed; and quantitative and qualitative evaluation of the items ensued. Phase III was the pilot testing phase where the scale was tested for internal-consistency reliability and inter-rater reliability. After pilot testing, the scale was reduced to two factors: Main Idea (which had 3 items) and Integration (which had 5 items). The resulting scale was reliable
and the factor loadings were appropriate based on the statistical and theoretical factor structure. Phase IV was the validation phase. The scale was evaluated for face validity, criterion related validity, Messik’s validity, and construct validity. Taken together the scale is reliable and valid for its stated purpose of measuring imagery.
Chapter Five: Imagery Assessment Results

During the pilot test, it was revealed that the Willard Imagery Observation Scales are reliable and valid for measuring imagery in children with and without autism. Two factors underlying imagery in autism were identified: Main Idea and Integration. All of the 8 retained items load consistently onto these factors based on the results of the Exploratory Factor Analysis. Next, it was possible to use the instrument to score the imagery skills of the participants. Results allowed for the comparison of the autism group to the typical group. This analysis answers, Research Question #2: Do children with autism show different abilities than typically developing children in terms of imagery-related integrative-comprehension skills as assessed by the imagery scale?

For this analysis, all of the results from the entire sample were included (N=71) (See Figure 8). The final sample included 23 children who were Typically developing and 48 children with an Autism Spectrum Disorder. There were more girls in the Typical group than the ASD group. The ASD group, on average had an IQ score that was 12 points lower than the Typical group. The groups were well matched on demographic variables. A description of the sample was provided in Chapter 3: Method.
Each participant was scored on their integrative-comprehension imagery skills using the WIOS. Results are shown in Table 11. A visual inspection of the descriptive statistics reveals that the autism group struggled more, on average, with these tasks as compared to the typical group. As shown, the autism group scored lower on every item and on the domain scores. Interestingly, the Typical group was especially proficient on the Main Idea domain; there was more variability in scores on the Integration domain.

Figure 8

*Final sample*
Table 11

Assessment Results

<table>
<thead>
<tr>
<th>Scores</th>
<th>Autism group Mean (SD)</th>
<th>Typical group Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=71</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Domain Scores</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WIOS Total (0-16)</td>
<td>9 (3)</td>
<td>13 (2)</td>
</tr>
<tr>
<td>Main Idea Domain (0-6)</td>
<td>4 (1)</td>
<td>6 (.2)</td>
</tr>
<tr>
<td>Integration Domain (0-10)</td>
<td>4 (2)</td>
<td>7 (2)</td>
</tr>
<tr>
<td>Global Coherence (0-100)</td>
<td>58 (26)</td>
<td>89 (9)</td>
</tr>
<tr>
<td><strong>Main Idea Items (0-2)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Characters</td>
<td>1.5 (.7)</td>
<td>2 (0)</td>
</tr>
<tr>
<td>2. Actions</td>
<td>1.8 (.4)</td>
<td>2 (.2)</td>
</tr>
<tr>
<td>3. Sequence</td>
<td>1.1 (.8)</td>
<td>2 (0)</td>
</tr>
<tr>
<td><strong>Integration Items (0-2)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Emotions</td>
<td>.7 (.8)</td>
<td>1 (.7)</td>
</tr>
<tr>
<td>5. Gestures</td>
<td>1 (.8)</td>
<td>1.4 (.5)</td>
</tr>
<tr>
<td>6. Causal-Connections</td>
<td>.7 (.8)</td>
<td>1.3 (.8)</td>
</tr>
<tr>
<td>7. Perspective Taking</td>
<td>1.3 (.7)</td>
<td>1.8 (.5)</td>
</tr>
<tr>
<td>8. Integrative Statements</td>
<td>1 (.7)</td>
<td>1.6 (.5)</td>
</tr>
</tbody>
</table>

*range of possible scores*

Assessment results indicate that the children with an Autism Spectrum Disorder scored, on average, 4 points lower on the WIOS total. The Typical group scored 2 or 3 points higher on the Main Idea and Integration domains, respectively. The Global Coherence rating was a subjective assessment by the examiner as to whether the child’s narrative made sense overall. As shown, the Autism group scored lower on Global Coherence, overall. Perhaps even more interesting is the amount of deviation in the Global Coherence ratings of the Autism group. There appeared to be a high degree of variability in the participants with regard to the clarity and coherence of their narratives. Overall, there appears to be a general trend indicating that the Typical group performed higher than the Autism group.
As shown in Table 11, there appears to be a difference in scores between children with ASD’s compared to typically developing. However, is that relationship statically significant? That is, do children with autism show significantly lower scores on the WIOS? In order to test this, a multiple regression will be used. However, prior to this analysis, it is generally important to test for correlation. As discussed in the Method section, the causal-comparative design is strengthened in this study because two of the three conditions of causality; correlation and isolation, were addressed. Although the third condition of temporal precedence could not be met, the predictive power of this research design was enhanced through the isolation of variables with statistical controls; and with the testing of correlations. As such, the first step is to conduct a correlation to see if a relationship exists between the constructs that are expected to correlate based on the researcher’s theory.

For this correlation analysis, the relationship of interest was: Diagnosis with WIOS Total score. It was predicted based on the theory that children with autism have imagery weaknesses, that this relationship would be positive and significant. That is, a child with Autism (coded with a 1) would score lower than a child with an ASD (Asperger’s or PDD NOS are coded with a 2); and the Typically developing children would score the highest of all (Coded with a 3). Thus, a positive and significant relationship would indicate that an increase in scores is associated with a Typical diagnosis; or that a lower score is associated with an Autism Diagnosis. The results of that analysis are provided in Table 12.
Table 12

Correlation between Diagnosis and WIOS Total

<table>
<thead>
<tr>
<th>Scores</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=71</td>
<td></td>
</tr>
<tr>
<td>WIOS Total</td>
<td>.70***</td>
</tr>
<tr>
<td>Domain Scores</td>
<td></td>
</tr>
<tr>
<td>Main Idea Domain</td>
<td>.60***</td>
</tr>
<tr>
<td>Integration Domain</td>
<td>.63***</td>
</tr>
<tr>
<td>Global Coherence Rating</td>
<td>.69**</td>
</tr>
<tr>
<td>Main Idea Items</td>
<td></td>
</tr>
<tr>
<td>1. Characters</td>
<td>.49***</td>
</tr>
<tr>
<td>2. Actions</td>
<td>.20</td>
</tr>
<tr>
<td>3. Sequence</td>
<td>.64***</td>
</tr>
<tr>
<td>Integration Items</td>
<td></td>
</tr>
<tr>
<td>4. Emotions</td>
<td>.33**</td>
</tr>
<tr>
<td>5. Gestures</td>
<td>.33**</td>
</tr>
<tr>
<td>6. Causal-Connections</td>
<td>.40***</td>
</tr>
<tr>
<td>7. Perspective Taking</td>
<td>.48***</td>
</tr>
<tr>
<td>8. Integrative Statements</td>
<td>.49***</td>
</tr>
</tbody>
</table>

Diagnosis bears a significant relationship with scores on the WIOS. Diagnosis is significantly associated with scores on the: WIOS Total, the Main Idea domain, and the Integration Domain. The relationships between diagnosis and item scores on the WIOS items are significant, with the exception of item 2: actions. Generally, a typical diagnosis was associated with the highest scores on the WIOS, indicating that the typically developing children in this sample told significantly better (more imagery-rich) narratives as compared to those with an Autism Spectrum Disorder. However, even though there is a significant relationship demonstrated here, will that relationship remain significant when considering covariates? Can the effect of diagnosis be isolated from the other
potentially important factors in WIOS scores? That is, can a diagnosis of autism predict lower scores on the WIOS, controlling for IQ, Age, and Gender?

**Results of regression on WIOS Total**

*Testing of assumptions*

Prior to running analysis, the assumptions of multivariate regression were tested. First, to check for *independence*, the unstandardized residuals were plotted against the standardized predicted values in order to assure that the error term is not related to the variance of the variables. A scatterplot of residuals was displayed and found to be randomly distributed around a mean of 0; thus the assumption of independence was indeed met. The assumption of *homoscedasticity* was tested by ensuring that the residual $r^2$ was around 0; and there were not a high concentration of residuals clustering above or below zero. Thus, these data are homoscedastic. Next, tests for *normality* were conducted, revealing a roughly normal distribution of the residuals. Finally, the assumption of the *error mean of 0* was assessed through descriptive statistics. Indeed, the mean of the unstandardized residuals was 0. Therefore, the assumptions of independence, homoscedasticity, normality, and error mean of 0, were all met.

Now that assumptions were met, a multivariate regression analysis could be conducted. This analysis could answer question #2: **Do children with autism show different abilities than typically developing children in terms of imagery-related integrative-comprehension skills as assessed by the imagery scale?** The dependent variable was score on the WIOS total. The independent variables were Diagnosis, Gender, age, and IQ. The results were as follows in table 13.
Table 13

Regression of Diagnosis on WIOS Total Score

<table>
<thead>
<tr>
<th>Model Summary</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n=71)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQ</td>
<td>.07**(.02)</td>
<td>.08(.02)**</td>
<td>.08***(.02)</td>
<td>.02(.02)</td>
</tr>
<tr>
<td>Age</td>
<td>.02(.02)</td>
<td>.03(.02)</td>
<td>.03*(.01)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td>2.5**(.84)</td>
<td>.00(.77)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Diagnosis

Autism=1, ASD=2
Typical=3

R squared | .13 | .16 | .26 | .54 |
F test    | 9.93** | 6.26** | 7.74*** | 19.34*** |

*p<.05, **p<.01, ***p<.001, (Standard errors)

Beta weights are unstandardized estimates

The answer to Research Question #2 is ‘yes.’ Regression results indicate that
while IQ, Age, and Gender play a role in scores on the Willard Imagery Observation
Scales; Diagnosis is the most significant predictor. That is, children with autism score
significantly lower than children with other ASD’s. Typically developing children score
higher than both groups on the WIOS. A diagnosis of autism seems plays a significant
role in predicting a child’s imagery ability as measured by this instrument (p<.001).

Model 4 was a test to see whether or not the effect of the control variables of IQ,
Age, and Gender would be lessened or removed when diagnosis was added to the model.
The effect of IQ, age, gender, together with diagnosis, predicted 54% of the variance in
WIOS total score. Model 4 was the best fitting model, explaining more of the variance in
than any other model (r^2=.54; F=19.34, p<.001). Therefore, the diagnosis variable was
the most significant predictor of scores on the WIOS total.
As predicted, the effects of the IQ and gender variables were rendered insignificant with the addition of the Diagnosis variable. That is, diagnosis explained much more of the variance in WIOS scores, such that IQ and age were no longer significant predictors. This finding is consistent with the hypothesis that children with autism think differently, rather than simply displaying lower cognitive abilities overall. This is a critical result, consistent with past research in autism; which provides that the autistic brain shows less proficiency than the typical brain in the coordinated use of imagery and language (Just, in press; Just, Cherkassky, Keller & Minshew, 2004).

Interestingly, the effect of Age was significant in model 4 (p<.05); although, not at the high level seen for the Diagnosis variable (p<.001). This means that Diagnosis and Age were covariates, together predicting higher scores on the WIOS. Older children scored a bit higher on the WIOS than younger children. It seems that maturation plays a role in imagery abilities; although not as significant of a role as diagnosis.

Overall, children with autism scored significantly lower on the WIOS total, relative to those with other ASD’s, and those who were typically developing. Taken together, results of this study reveal some significant differences, not in overall cognition, but in the way the children with autism think. This important difference has a myriad of implications for practice which will be discussed in Chapter 6: Discussion.

The next issue to address in this study was whether or not the children with autism showed significant performance differences on the domain scores: Main Idea and Integration. This analysis would provide researchers with information about the component skills of imagery that may be impaired in the autism population.
The next step was to run a multi-variate regression in order to test the significance of the relationship between a diagnosis of autism and lower scores on the Main Idea domain. This analysis would answer: **Research Question #2a. Do children with autism show more difficulty with understanding and explaining the Main Idea of a story as measured by the Main Idea domain? (Testing the Weak Central Coherence Theory)**

In order to answer this question, a multi-variate regression was run on the entire sample (N=71). This analysis would answer whether or not Diagnosis was a significant predictor of scores on the Main Idea Domain; controlling for the effects of IQ, age, and gender. The results of this analysis are provided in Table 14.
Table 14

*Regression of Diagnosis on the Main Idea Domain (N=71)*

<table>
<thead>
<tr>
<th>Model Summary (n)</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAIN IDEA</td>
<td>DOMAIN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQ</td>
<td>.03** (.01)</td>
<td>.03** (.01)</td>
<td>.03** (.01)</td>
<td>.01 (.01)</td>
</tr>
<tr>
<td>Age</td>
<td>.01 (.01)</td>
<td>.01 (.01)</td>
<td>.01 (.01)</td>
<td>.01 (.01)</td>
</tr>
<tr>
<td>Gender</td>
<td>.97** (.34)</td>
<td>.14 (.35)</td>
<td>.95*** (.21)</td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td>R squared</td>
<td>.14</td>
<td>.23</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td>F test</td>
<td>9.26**</td>
<td>5.36**</td>
<td>6.63**</td>
<td>11.94***</td>
</tr>
</tbody>
</table>

1. Characters

| IQ                | .01*(.00)           | .01*(.00)           | .01 (.00)           | .00 (.00)           |
| Age               | .00 (.00)           | .00 (.00)           | .00 (.00)           | .00 (.00)           |
| Gender            | .34*(.16)           | .03 (.17)           | .35** (.10)         |                     |
| Diagnosis         |                     |                     |                     | .07                 |
| R squared         | .07                 | .13                 | .26                 |                     |
| F test            | 4.99*               | 2.61                | 3.3*                | 5.92***             |

2. Actions

| IQ                | .00 (.00)           | .00 (.00)           | .00 (.00)           | .00 (.00)           |
| Age               | .00 (.00)           | .00 (.00)           | .00 (.00)           | .00 (.00)           |
| Gender            | .12 (.01)           | .04 (.11)           | .09 (.06)           |                     |
| Diagnosis         |                     |                     |                     | .00                 |
| R squared         | .00                 | .04                 | .03                 | .05                 |
| F test            | .00                 | .05                 | .59                 | .92                 |

3. Sequence

| IQ                | .02** (.01)         | .02** (.01)         | .02** (.01)         | .01 (.01)           |
| Age               | .00 (.00)           | .00 (.00)           | .00 (.00)           | .00 (.00)           |
| Gender            | .41 (.20)           | -.13 (.20)          |                    |                     |
| Diagnosis         |                     |                     | .61*** (.11)        |                     |
| R squared         | .16                 | .16                 | .21                 | .45                 |
| F test            | 13.27**             | 6.55**              | 6.02**              | 13.57***            |

*p<.05, **p<.01, ***p<.001, (Standard errors)

Beta weights are unstandardized estimates
Results of regression on Main Idea Domain

The answer to the research question is, ‘yes.’ The children with autism showed significantly lower performance on the Main Idea domain. In Model 4, the variables of IQ, Age, and Gender are all rendered insignificant with the addition of the diagnosis variable. Diagnosis is a significant predictor of scores on the Main Idea domain (p<.001). Model 4 is the best fitting model, explaining 42% of the variance in scores on the Main Idea domain ($r^2=.42; F=11.94; p<.001$). This means that although the variables of IQ, Age, and Gender have an effect on scores, diagnosis explains more of the variance in scores, causing the effect of these other factors to be insignificant. This finding is consistent with the hypothesis that children with autism or an ASD, would show weaker abilities to grasp and describe the Main Idea of a story, controlling for the effects of IQ, Age, and Gender. This finding is consistent with past research on the Central Coherence Account (Diehl et al., 2006; Nuske & Baven, 2010) indicating that children with autism tend to struggle to understand the ‘gist’ of a story.

Items. Similar to the results of the overall domain score, scores on the characters item were predicted by diagnosis ($r^2=.26; F=5.92; p<.001$). This finding is expected based on past research indicating that children with autism tend to have difficulty describing characters (Rutherford, Young, Hepburn & Rogers, 2007; White, Hill, Happe, & Frith, 2009). The results of the Actions (item 2) item are puzzling. None of the models explain a significant amount of variance in scores on this item. It appears that either the item is not well written or defined; or that there are other variables that explain it which are not included in the model. Diagnosis was also the most significant predictor of scores
on the sequence item (p<.001), controlling for the effect of IQ, Age, and Gender ($r^2=.45$; $\text{F}=13.57; p<.001$). This finding is predicted based on past research indicating that children with autism struggle to understand sequencing concepts (Philofsky, Fidler & Hepburn, 2007).

Now that the question of whether or not the Main Idea domain is impaired in autism has been considered, it is time to consider the Integration domain. Thus, this analysis would answer, **Research Question #2b. Do children with autism show more difficulty integrating story events as measured by the Integration domain of the imagery scale? (Testing the ‘functional under-connectivity’ theory)**

To answer this question, a multi-variate regression was conducted. The regression tested whether or not Diagnosis was a significant predictor of scores on the Integration Domain, controlling for the effects of IQ, Age, and Gender. The results were as follows in Table 15.
Table 15
Regression of Diagnosis on the Integration Domain

<table>
<thead>
<tr>
<th>Model Summary</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n=71)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>INTEGRATION DOMAIN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQ</td>
<td>.04*(.02)</td>
<td>.05**(.02)</td>
<td>.05**(.02)</td>
<td>.01(.02)</td>
</tr>
<tr>
<td>Age</td>
<td>.02(.01)</td>
<td>.02(.01)</td>
<td>.02(.01)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>-1.4 (.60)</td>
<td>-1.4(.60)</td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
<td>1.93***(.36)</td>
<td></td>
</tr>
<tr>
<td>R squared</td>
<td>.09</td>
<td>.12</td>
<td>.19</td>
<td>.43</td>
</tr>
<tr>
<td>F test</td>
<td>6.78*</td>
<td>4.48*</td>
<td>5.23**</td>
<td>12.58***</td>
</tr>
</tbody>
</table>

4. Emotions

| IQ            | .01(.01)  | .01(.01)  | .19(.01)  | .00(.00) |
| Age           | .01(.00)  | .00(.00)  | .01(.00)  |          |
| Gender        | .66**(.21) | .47(.24) |          |          |
| Diagnosis     |          |          | .22(.14) |          |
| R squared     | .01      | .07      | .19      | .22      |
| F test        | .96      | 2.39     | 5.14**   | 4.56**   |

5. Gestures

| IQ            | .00(.01)  | .01(.01)  | .01(.01)  | .00(.01) |
| Age           | .00(.00)  | .00(.00)  | .00(.00)  |          |
| Gender        | .28(.21)  | - .03(.24) |          |          |
| Diagnosis     |          |          | .36*(.14) |          |
| R squared     | .01      | .02      | .05      | .13      |
| F test        | .51      | .62      | .78      | 2.53*    |

6. Causal Connections

| IQ            | .01(.01)  | .01(.06)  | .01(.01)  | .00(.00) |
| Age           | .00(.00)  | .00(.00)  | .00(.00)  |          |
| Gender        | .01(.23)  | - .44(.23) |          |          |
| Diagnosis     |          |          | .51**(.14) |          |
| R squared     | .04      | .05      | .05      | .20      |
| F test        | 3.08     | 1.64     | 1.08     | 4.18**   |

*p<.05, **p<.01, ***p<.001, (Standard errors)
Beta weights are unstandardized estimates
Table 15 (continued)

*Regression of Diagnosis on the Integration Domain (N=71)*

<table>
<thead>
<tr>
<th>Model Summary (n=71)</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>7. Perspective Taking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQ</td>
<td>.02**(.00)</td>
<td>.02***(.00)</td>
<td>.02***(.00)</td>
<td>.01*(.01)</td>
</tr>
<tr>
<td>Age</td>
<td>.01(.00)</td>
<td>.01(.00)</td>
<td>.01(.00)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>.30(.17)</td>
<td>.00(.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
<td>.33**(.11)</td>
<td></td>
</tr>
<tr>
<td>R squared</td>
<td>.16</td>
<td>.19</td>
<td>.22</td>
<td>.31</td>
</tr>
<tr>
<td>F – test</td>
<td>12.7**</td>
<td>7.9**</td>
<td>6.38**</td>
<td>7.56***</td>
</tr>
<tr>
<td><strong>8. Integrative Statements</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IQ</td>
<td>.01(.01)</td>
<td>.01(.01)</td>
<td>.01(.01)</td>
<td>.00(.01)</td>
</tr>
<tr>
<td>Age</td>
<td>.00(.00)</td>
<td>.00(.00)</td>
<td>.01(.00)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>.40*(.18)</td>
<td>.05(.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
<td>.41**(.12)</td>
<td></td>
</tr>
<tr>
<td>R squared</td>
<td>.04</td>
<td>.05</td>
<td>.12</td>
<td>.26</td>
</tr>
<tr>
<td>F test</td>
<td>2.54</td>
<td>1.93</td>
<td>3.0*</td>
<td>5.83***</td>
</tr>
</tbody>
</table>

*p<.05, **p<.01, ***p<.001, (Standard errors)

Beta weights are unstandardized estimates

**Results of Regression on Integration Domain**

The answer to the research question is ‘yes.’ Results of the regression of diagnosis on the Integration domain indicate that diagnosis is a significant predictor of scores. In Model 4, the effect of IQ, Age, and Gender are all rendered insignificant with the addition of Diagnosis to the model. This result was hypothesized. It was hypothesized that Diagnosis would be the most significant predictor of scores on the Integration Domain, controlling for the effects of IQ, Age, and Gender. Indeed, Model 4 was the best fitting model, explaining 43% of the variance in scores on the Integration domain. ($r^2=.43$, 12.58, p<.001). This result is consistent with the theory that children with autism
display a ‘functional under-connectivity’ in the brain’s ability to integrate across the structures that process language and those that process images (Just, Cherkassky, Keller, & Minshew, 2004; Just, in press.).

**Items.** Results of the regression of diagnosis on the *emotions* item reveal that IQ, Age, and Diagnosis were not a significant predictor of scores. These results were not predicted (see Discussion for an explanation of findings from item scores). Results of the regression of diagnosis on the *gestures* item are that Diagnosis is a significant predictor ($r^2=.13; F=2.53, p<.05$). This is clearly an anticipated result because limited gesture use is known to be a common impairment in children with autism (Philofsky, Fidler, & Hepburn, 2007). However, this result is especially interesting in this study because past research has identified limited gesture use as a sign of imagery problems (Bell & Bonetti, 2006). Bell and Bonetti explain, “Many children…do not use gestures when they talk, probably as a good indication as any that they are also not imaging vividly” (p.27). Implications of these findings are provided in Chapter 6: Discussion.

Results of the regression of on the *causal connections* item reveals that Diagnosis is the only significant predictor ($r^2=.20; F=4.18, p<.01$). This effect of diagnosis on causal connections was anticipated based on past research (Diehl, Benetto, & Young, 2006).

Results of the regression on the *perspective taking* item reveals that Model 4 is the best fitting model, which included the Diagnosis variable ($r^2=.31; 7.56, p<.001$). Thus, although IQ was a significant predictor of scores on the perspective taking item, Diagnosis is the most significant. This was expected based on past research indicating
that children with autism have impaired Theory of Mind (Rutherford et al., 2007).

Finally, for the *Integrative Statement* item, Diagnosis is the only significant predictor of scores ($r^2=.26; F=5.83; p<.001$). Taken together, diagnosis predicts scores on all of the domains, and most of the items; controlling for the effects of IQ, Age, and Gender.
Chapter Six: Discussion

A sample of 71 children with (N=48) and without (N=26) an autism spectrum disorder was compared in their use of imagery. Results were that children with autism score significantly lower on measures of imagery as compared to typically developing children. This effect holds up while controlling for the effects of IQ, Gender, and Age. These results have important implications. This chapter begins with the answers to the research questions regarding imagery in autism. Next, is a discussion of the implications for these findings for research and practice. Then, there is a discussion of the limitations of the study. Finally, this chapter concludes with the contributions of this research.

The first research question revolved around scale development. A literature review revealed that while several measures of imagery exist, none of these were appropriate for this study because the instruments were generally developed for adults; never tested in clinical populations (such as autism); and did not measure integrative-comprehension imagery. Thus, a new scale was developed. Results answered Research question #1: Can a valid and reliable scale be developed to assess integrative-comprehension skills in children with and without autism?

The answer to question #1 was yes. The Willard Imagery Observation Scales demonstrated adequate reliability. Inter-rater reliability was demonstrated for the Total score (ICC=.93); for the Main Idea domain (ICC=.94) and the Integration domain (ICC=.86). The internal consistency of the instrument was adequate as well (α=.72). On
the test of the final scale (Version 3), there were no problematic items. Thus, all of the domains and items of the WIOS showed adequate reliability.

The scale was assessed with regard to the factor loadings of each domain. Indeed, all of the items loaded on the theoretical factors. All of the Main Idea domain items had factor loadings $\geq .30$ (characters=.89; actions=.64; sequence =.75). All of the Integration items loaded consistently onto the Integration domain; demonstrating factor loadings $\geq .30$ (emotions=.69; gestures=.30; causal connections=.74; perspective taking=.79; and integrative statements =.58). Thus, the factor structure has demonstrated validity in terms of capturing the constructs that were intended to be measured by the items.

**Validity.**

The scale held-up as consistently reliable in Pilot 3 (N=25) and in the Validation Sample (N=46). Thus, the scale seems to demonstrate a degree of generalizability. Face Validity was tested through rigorous evaluation by a Subject Matter Expert Panel. Content Validity was demonstrated through systematically narrowing down the items and test domains based on a quantitative and substantive evaluation. Messick’s Validity was demonstrated because the test results are highly likely to be helpful to examinees. Criterion-related validity was not tested because there were not enough data on standardized comprehension instruments that could be tested for correlation with the WIOS results. Construct Validity was assessed through a correlation between Diagnosis and WIOS Total score, which was high and significant ($r=.70; p>.001$). Taken together, the WIOS demonstrated strong reliability and preliminary evidence for validity as a measure of imagery.
Scale development results are similar to past research indicating that a valid and reliable scale can be developed (Isaac & Marks, 1994); however, the task of measuring imagery can be “vexing” (Gambrel & Bales, 1985, p.) and “puzzling” (Bell, 1991b, p.24). Most imagery measurement studies ended with a statement about how complex and challenging it is to measure imagery (Sadoski, 1983). However, researchers point to some of the commonalities between imagery measurement tools as promising indicators that researchers are coming closer to capturing this complex construct (Sadoski, 1983). One such commonality, uncovered in this study, was the identified domains. While there is a fair amount of debate over the domains of imagery, virtually every imagery measure included some approximation of a Main Idea construct and an Integration construct. Taken together, development of an imagery measure was occasionally formidable but always worthwhile.

An important take-away from the scale development process was the idea that the scale could be used for children from various diagnostic categories (Autism, ASD, Typical). The researcher originally thought the Typical group was only included as a basis for comparison (a control group). However, during the research process, an interesting result was uncovered: this scale can be just as effective for Typically developing children. The researcher learned while developing the scale that it was more important to identify the skills required for optimal outcomes, rather than simply focusing on deficits. In so doing, the model that was developed (Figure 3: Theoretical Model of Integrative-Comprehension Imagery) is applicable to all readers, not only children with autism.
However, one of the significant strengths of this study was the clinical population included in the standardization sample. A sample of this nature with so many clinically diagnosed children on the autism spectrum is rare; even in studies conducted by well-known researchers, published in respected journals. Thus, the clinical sample included in this study strengthened this instrument’s validity substantially. The instrument was designed such that any child who could tell a story could participate. The story included no words, so reading level was not a factor. Further, the children ranged in IQ levels from 80-142. Thus, the scale seems to demonstrate a unique kind of clinical validity. The scale is sensitive to slight, but significant, cognitive differences in children with autism as compared to typically developing. The scale is ‘inclusive’ in that children with various cognitive abilities can participate. As a result, this scale has a wide range of clinical and educational applications.

**Imagery Differences in Autism.**

Once the scale was shown to be reliable and valid, the instrument was used to assess the imagery abilities of children with and without autism. The researcher wanted to see whether or not the hypothesized weakness in imagery skills is apparent in the autism sample, based on the results of this test.

In order to analyze the assessment results and begin to understand imagery in autism, multivariate regressions were run on the WIOS Total score. Thus, it was possible to answer: **Research question #2: Do children with autism show different abilities than typically developing children in terms of imagery-related integrative-comprehension skills as assessed by the imagery scale?**
The answer to research question #2 is yes. Diagnosis was the most significant predictor of scores on the WIOS, controlling for the effects of IQ, Age, and Gender. The correlation between diagnosis and scores on the WIOS was $.70 \ (r=.70; \ p<.001)$. This means that 70% of the variance in scores on the WIOS can be explained by diagnostic status. In the regression analysis on the WIOS total, 54% of the variance in scores was explained by Diagnosis, while controlling for the effects of IQ, Age, and Gender \( (r^2=.54; \ F=19.34, \ p<.001) \). The covariates of IQ, Age, and Gender showed up as significant in some of the models. However, this effect was generally rendered insignificant with the addition of the diagnosis variable.

One of the most important findings of this study was that IQ was not the most significant predictor of imagery skills. This is critical information for interventionists. If the cognitive differences associated with autism were simply quantitative (lower IQ), there is a sense that a ceiling exists as to possibilities for intervention. Children with lower IQ’s have difficulties with a wide variety of tasks; which can be discouraging. In this study, results suggest that there is a qualitative difference in the cognitive abilities of children with autism. Perhaps, the brain has a great deal of untapped potential, but is not as well integrated (the various functions are not talking to each other). If this is the case, these results offer up a myriad of opportunities for intervention to teachers and interventionists (See Implications of Practice section).

The Age variable generally did not play a significant role in scores on the WIOS. On all of the domain scores and most item scores, Age was a non-significant covariate. Even when the Age variable was only combined with IQ in predicting WIOS scores,
there was no significant effect. This is somewhat surprising in that one might think maturation would play a role based on imagery research indicating that these skills develop over time (Pressley, 1976; Pressley, 1977; Pressley & Levin, 1978; Sadoski, 1983).

However, there could be two reasonable explanations for this result. First, the range of ages was restricted to children who were in the elementary school grades, where imagery is generally in place for them (Pressley, 1976; Sadoski, 1983). Secondly, the age variable is impacted by the fact that most of the children in this study have a neurodevelopmental disability. As such, the children’s skills may actually decline in comparison to peers over time (the gap widens with age). Thus, it may be that the combination of typical maturation in the Typical sample, and declining skills with age in the ASD population, may have essentially washed out the effect of age.

The Gender variable was a significant predictor of scores on the WIOS in some of the models but Diagnosis was still more significant. For example, the Gender variable was significant on the Main Idea domain, the characters item, and the emotions item. Research shows that there are gender differences in imagery abilities (Isaac & Marks, 1983), so this might be expected to be a significant covariate. While this result might be expected in some ways, it also may be a partially spurious relationship. It is possible that the gender variable was impacted by sample selection. There were many more girls in the Typical sample as compared to the Autism sample and this may have confounded results. Further, as soon as diagnosis was added to the model (in Model 4), the effect of Gender was rendered non-significant (See Limitations section).
In all of the Models for almost every item, the Diagnosis variable was the most significant variable. This finding upholds the central hypothesis of this study. That is, the researcher believed children with autism would show this significantly different way of processing and re-telling stories in spite of the potentially important covariates of IQ, Age, and Gender. Children with autism, in this sample, seem to indeed show this qualitative difference in the way they think; showing a difficulty using words to describe images. This begs the question as to whether or not some of the speech and language problems evidenced in this population have a foundation in imagery. Based on past research, there are several evidence-based interventions that could address this problem (Bell, 1991b; Gambrell & Bales, 1986; Joffe, Cain, & Maric, 2007; Pressley, 1976). Thus, it is possible that if children with autism received early intervention for imagery, their communication patterns may be improved and enhanced.

**The Weak Central Coherence Account of autism.**

Next, this study intended to uncover whether or not the children with autism in this study scored statistically significantly different on the component skills of imagery as measured by the domains. First, it was interesting to assess whether or not the children with autism displayed consistently lower scores on the Main Idea domain. This would answer, **Research Question #2a: Do you children with autism show more difficulty in understanding the main idea of a story as measured by the Main Idea domain of the imagery scale? (Testing the Central Coherence Theory)**

The answer to research question is ‘yes.’ Diagnosis was the most significant predictor of scores on the Main Idea Domain (p<.001). This means that when children with autism
told stories from a picture book, they struggled to identify or describe the ‘gist’ or the central idea. Narratives without a main idea lack coherence and clarity (Diehl, Bennetto, & Young, 2010). A person who is listening to a narrative without a main idea would likely ask questions like, “okay, but what basically happened?” or “can you back up and tell me what we are talking about?”

This finding is quite meaningful to the study of autism. First, while the Weak Central Coherence Account of autism has been under debate in recent studies (Tager-Flusberg, 2006), the theory is consistent with these results. It is well understood that children with autism tend to concentrate on details. For example, if a child with autism was playing with a toy train, he may become fascinated with the wheels; or when playing with a doll, he may focus on the eyes. It seems that they often miss the ‘gestalt’ of an object. Rutherford, Young, Hepburn & Rogers (2007) proposed that this tendency to play with toys in a non-functional manner demonstrates an overall weaker understanding of symbolic representations. This means, they may indeed not understand that the doll is a symbolic toy for a real baby or that a toy train is meant to represent a real train. This research also indicates that difficulties with symbolic representation in play may have a basis in the autism symptom of weak Theory of Mind (ToM) (White, Happe, & Frith, 2009). Thus, weaker symbolic representations, weaknesses in Theory of Mind, and extreme focus on details, all seem to converge on a common theme. These apparently idiosyncratic symptoms of autism seem may all center around the Weak Central Coherence Account.
Further interesting support for this idea came from researchers Diehl, Bennetto & Young (2006) who found that children with autism could sometimes answer questions about the main idea but they failed to use the main idea or to structure their stories around a central concept. In the current study, consistent with this research; even when the children with autism had adequate knowledge of the main idea, their narratives were still jumbled and disjointed. This research is consistent with the work of researchers Philofsky, Fidler & Hepburn (2007) who describe this ‘lack of cohesion in discourse’ as a key language symptom seen in the autism population.

In this study, it was interesting, qualitatively, to watch the way children with Asperger’s told stories. Although their narratives included most of the story elements, the stories were generally laden with irrelevant details. For example, one child with Asperger’s was describing a turtle sitting on a log. She immediately launched into a discourse about the nocturnal and diurnal turtle species, and into whether or not Charles Darwin is the father of the theory of survival-of-the-fittest! Thus, even though she provided an interesting narrative, it lacked a general coherence because she got lost in details and failed to center her story around the main idea; which is consistent with past research (Diehl et al., 2006).

Overall, the results of this study are consistent with the Weak Central Coherence account of autism (Nuske & Baven, 2010; Diehl et al., 2006). The children in this study seemed to show many of the same struggles evidenced in previous research. Further support for this specific and qualitative difference is found in the fact that diagnosis remains significant even when controlling for IQ. Many of the children with autism or
Asperger’s Disorder in this study had IQ’s that were well above average; some had IQ’s in the superior range. Still, they struggled to describe the main idea of stories.

This ability to describe the main idea is important not only for reading comprehension but also for communication and social skills (Bell & Bonetti, 2006; Mather & Goldstein, 2008). For example, Diagnosis was a significant predictor of scores on the characters and sequence items of the Main Idea domain. This means that the children with autism had trouble identifying characters and describing a sequence of events in their stories. These findings are consistent with past research on the weaker Narrative Production abilities of children with autism (Diehl, Benetto & Young, 2006). Diehl, Benetto & Young have shown that children with do not tell coherent stories. Philofsky, Fidler, and Hepburn have demonstrated that a lack of coherence in narrative discourse is a common symptom of autism (Philofsky, Fidler & Hepburn, 2007).

It seems clear that these deficit areas would significantly impact conversation skills and social communication (Bell & Bonetto, 2006; Gambrell & Bales, 1985; Mather & Goldstein, 2008). Mather and Goldstein (2008), explain, that weaknesses in imagery can manifest into, “poor social skills; difficulty paying attention [and] extreme difficulty with reading comprehension” (319). It is likely that if a child with autism tells an incoherent story or communicates in a jumbled and disjointed fashion, peers may become bored or frustrated. The research here provides initial support for the theory that imagery may underlie some of these challenges. The good news about this is that some of these deficits may be amenable to intervention (See the Implications for Practice section).

The next component skill of integrative-comprehension imagery that was tested in this study is integration. That is, the results of the Integration domain provided critical insights into the imagery abilities of children with and without autism. These results answered, Research Question 2b. Do children with autism show more difficulty integrating story events and ideas as measured by the Integration domain of the imagery scale? (Testing the ‘functional under-connectivity’ theory)

The answer to research question #2b is ‘yes.’ A diagnosis of autism or an Autism Spectrum Disorder was the only significant predictor of scores on the Integration Domain; controlling for the effects of IQ, Age, and Gender (p<.001). This means that as the children with autism told stories, there was a lack of connection between story events and ideas. For example, on item 8: integrative-statements, typically developing children tended to provide a statement that connected the whole story together. One child said, “Huh, that’s strange that frogs and pigs can fly in this story. What a weird book. I wish I could fly.” Thus, these children had the ability to tie together the whole story and integrate all that they have learned into a coherent whole. Diagnosis was the only significant predictor of scores on this integrative statement item, when controlling for IQ, Age, and Gender (p<.001). Thus, the children with autism in this sample tended to tell stories that were poorly connected or not well integrated.

This finding is significant for the science of imagery in autism. A recent body of literature suggests that the brain of autism shows a ‘functional under-connectivity’ (Just in press, Just et al., 2004, Kana et al., 2006). These researchers used a functional MRI to
test the activation rate in the brain while children and adults with autism were listening to imagery-rich stories. They found a lack of activation between the brain structures that process imagery and those that process language. Past research on imagery has revealed that children with imagery-rich narratives tend to show a ‘functional use of imagery…seen as interacting with the verbal processes reciprocally’ (Sadoski, 1983, p.121).

The functional under-connectivity theory is well vetted in the current study. Consider the task at hand on the WIOS. Children are asked to tell stories based on pictures. If there was a disconnect between the parts of the brain that process images and language, this task would be difficult. Indeed, study results suggest that the children with autism struggled significantly with the integration of images and language in their narratives; which suggests that there may be a functional underconnectivity in the brain of autism.

Regarding item scores, Diagnosis was also a significant predictor of several items within the Integration domain. Diagnosis was the only significant predictor of scores on the gestures item. This finding should not be surprising in that children with Autism Spectrum Disorders generally show limited gesture use. It is interesting, though, that gesture use is limited in children with poor imagery as well (Bell & Bonetti, 2006). Research shows that children with poor imagery do not use many gestures. Bell and Bonetti (2006) hypothesized that children with poor imagery do not use gestures because the images are not clear in their own minds.
Bell & Bonetti (2006) explain,

many children...do not use gestures when they talk, probably as good an indication as any that they are not imaging vividly” (p.27).

Thus, it may be that some of this limited gesture use in autism may have foundations in imagery.

This finding could have some important implications to the study of autism. It may be that children with autism are not less interested or less motivated to communicate, as some may think (although, of course there can be many exceptions to any ‘rule’ about autism). It may be that they are having trouble picturing the meaning conveyed by words; paired with a weaker ability to describe those images to others. Perhaps, if they can be taught to integrate their images with the language in the way neurotypical people do, their communication will improve and gesture use will increase. Indeed a qualitative review of the narratives used in this study, reveals more gesture use, generally, in the children who told richer, more engaging stories.

The final item in the Integration Domain where diagnosis was the only predictor of scores was on the causal connections item. This means that when the children told stories from a picture book, the children with autism had trouble explaining how events were causally connected, relative to the typical group. For example, a typically developing child might have said, ‘He got up to eat a sandwich because he felt hungry in the middle of the night.’ A child with autism said, ‘There, he is eating a vegetarian meat sandwich at night.’ Thus, although not factually inaccurate, the child with autism offered irrelevant information and failed to connect the events in the story into a logical causal...
sequence. Causal connections help the listener to get the ‘why’ of what has happened in the story. This finding could have important implications for practice. It is possible that children with autism could be explicitly taught to link together story elements into a causally connected, well sequenced narrative (See Implications for Practice).

Further interesting in these results was the finding that no other variable had a significant effect on causal connections. IQ, Age, and Gender were insignificant in all of the models regressed on causal connections. This is somewhat peculiar in that it might be expected that intelligence and maturation would predict some differences in making causal connections in stories. However, in this study, it was revealed that diagnosis was the only variable that predicted scores on this item. Results suggest that the qualitative brain differences in autism were to blame for impaired ability to causally connect story events.

Another item score in the Integration domain that was affected by Diagnosis was the perspective taking item. Perspective taking, though not always a diagnostic feature of the disorder, is considered a symptom of autism (Rutherford, Young, Hepburn, & Rogers, 2007). What is interesting here, though, is that this lack of perspective-taking ability may have roots in imagery, based on the results of past imagery assessment research (Sadoski, 1983). This result begs the question as to whether the children are not considering the mental state of the characters; or if they cannot picture the story or characters well. Or, perhaps, they can picture the characters and assess their mental states, but they cannot describe those images with words. That is, results uncover some cognitive differentiation that has not been considered in previous studies. Perhaps, the children with autism could
be taught to better image the characters, use a thought bubble to describe what the characters are thinking, and then to put those images into words. This may improve their communication overall.

The finding that children with autism struggle to take on the perspective of characters in stories has the potential for substantial academic problems. This lack of perspective-taking will assuredly impact reading comprehension. Generally some level of empathy or understanding of another’s thoughts or intentions is required when reading. Books about historical events, social studies, or fictional texts generally all include comprehension questions at the end of each section about what the characters are thinking or feeling. Perhaps, if children with autism could be taught to take on the mental perspective of characters, their reading comprehension would improve overall (See Implications for Practice).

**Implications for Practice**

The first implication for practice is that a valid and reliable imagery screener can be used to assess the imagery abilities of children with and without autism. The initial results of the WIOS are promising in that the instrument does appear to be uncover a significant weakness in imagery experienced in the autism population. It was discovered; during scale development that the WIOS, that the scale could be used for children who are Typically developing as well. The skills assessed on the instrument are important for all children. However, the instrument has special applications in the autism population because the test items were specifically chosen based on known autism symptoms. The
WIOS can, thus, be used as a needs-based screener that could pinpoint the literacy needs of children with and without autism.

This scale has made two significant contributions to educational practice: 1) children with comprehension problems can be screened for imagery deficits; 2) results of this screener indicate specific and targeted areas of need in children with autism. These targeted skill deficits can be used to inform instruction and practice. Rather than offering up yet another listing of skill deficits in this population, these results are targeted and shown in research to be amenable to intervention. As such, although this is certainly not an intervention study, the intent of the WIOS is to help practitioners uncover opportunities for instructional support or intervention. In order to have a justification for specific intervention, the results of the WIOS should be compared against a standardized measure of comprehension such as the Gray Oral Reading Test (GORT), the Woodcock Johnson Tests of Achievement, 3rd Edition, Normative Update (WJ-III, ACH, NU), or the Clinical Evaluation of Language Fundamentals (CELF).

As an initial screener; provided that the instrument is tested on larger populations in future studies (see implications for research) and that the test is correlated with the results of standardized tests, the WIOS could provide opportunities for targeted instructional techniques or interventions. First, clinicians could take a look at the WIOS total score. In this study, the scores ranged from 2-16, with a mean of 10 and a standard deviation of 3. Interventionists might consider any score between 2-7 (1-2 SD below the mean) as a potential indicator of imagery problems. These imagery problems could indicate that an evidence-based imagery intervention is indicated. There are a variety of
evidence-based programs available on the market such as the Lindamood-Bell
Visualizing-Verbalizing program; the Language! program, or Reading Recovery.

Experts in imagery interventions provide useful imagery tutoring techniques (Bell
& Bonetti, 2006). Bell & Bonetti suggest that imagery instruction should be direct and
explicit. They suggest that the imagery instruction is a dynamic interaction between the
teacher and student where questions are common; and imagery errors are teaching
opportunities. Their “error handling” (p.23) technique implies that teachers should:

1) make note of the student’s response, 2) find a spot in which to actively engage
him, and 3) help him analyze his response, and 4) and help him compare his
response to the stimulus. p.23

That is, imagery instruction involves an examination of the child’s images in attempt to
help the child enhance these images such that comprehension is improved.

Interventionists working with children who have impaired imagery can provide support
through explicit teaching and reciprocal interaction. This will allow children to learn
from the imagery instruction of the teacher as they are participants in their own learning.

Classroom-wide instructional supports could also be employed to address imagery
weaknesses. Imagery is essentially a strategy that is used for comprehension (Bell,
1991b; Pressley, 1977). Thus, best practice strategy instruction models could be used to
improve comprehension. (Strickland & Feeky, 1985). Strategy instruction involves:
modeling, guided practice, feedback, and independent performance (Reid, & Lienemann,
2006). The teacher would model his or her own use of imagery by describing the images
the story evokes in the teacher’s mind (Strickland & Feaky, 1985). Next, a teacher could
instruct students to use imagery with directions like: “Remember to make pictures in your head” or “It’s easier to remember what you read when you make a movie in your mind.”

Next, in a ‘guided practice’ or ‘supported practice’ phase (Reid & Lienemann, 2006), students would discuss their own images and assess their accuracy. For example, if the child has an image of a girl with 2 red balloons, and the teacher knows that the story included a bouquet of different colored balloons, the child could be instructed to correct the error. Finally, the children can be directed to use imagery independently. Best practice strategy instruction involves the gradual transfer of responsibility from the teacher to the student. Over time, students could construct their own mental models, and then check-in with teachers to see if they are on-track (Strickland et al., 1985). These interactive and dynamic approaches may be effective practices for teaching imagery.

Regarding intervention for deficits in the understanding of the Main Idea, it would be important for teachers to help children with autism identify the main characters in the story, define the order of events that occurred, and to make some general conclusions about the story. Further, past research has shown that even when children with autism know the main idea, they do not use it effectively in their own narratives (Diehl, Bennetto, & Young, 2006). Interventionists should consider teaching children to start with the main idea as they tell stories. Then, each story event should be told, in order, and linked back to that main idea.

Regarding interventions for Integration, children could be directly instructed in all of the skills assessed by this domain including: emotions, gestures, causal connections, perspective taking, and integrative statements. The causal-connection item results in this
study indicate that teachers should know it may be more difficult for children with autism
to see how events in a story are related to each other. For example, if the story read that
‘frogs ran into the clothes on the clothesline’ the child may not be able to explain why
clothes were on the ground in a later part of the story.

Visual inspection of data revealed that most children with autism did not provide
any causal connections. Some children with autism provided a loose causal connection
that was not well supported. For example, one child said, ‘he feels sad because of frogs.’
This child probably knows that there is a relationship between that emotion and an event
in the story but is somehow unable to articulate his knowledge. Thus, interventionists
working with children with autism on reading comprehension should not only ask them
about cause-and-effect relationships, but also provide concrete and explicit explanations.
Children with autism could benefit from modeling, direct-instruction, guided practice,
and independent practice to understand causal-connections.

Further evidence for intervention was uncovered on the perspective-taking and
emotions items. It is important for teachers to be aware that children with autism
generally do not understand what other people are thinking or feeling. In this study, the
children with autism had particular difficulty taking the mental state of another. These
findings are consistent with past research (White et al., 2010). For example, a typical
child might say, “This guy here is wondering why frogs are flying outside is window. He
thinks that’s strange!” A child with autism might be able to identify his physical
perspective. However, when prompted, these children with autism could not understand
what this man might be thinking. Thus, reading interventionists might consider acting out
stories, or guided role-play activities, in order to help children become more aware of how they might feel ‘in someone else’s shoes.’

Finally, interventionists should know that children with autism might have trouble putting the whole story together into a meaningful, coherent whole, as represented by the integrative statements on this scale. Visual inspection of the videos revealed that children with autism would simply close the book without comment; or would wrap it up with a simple, ‘The End.’ Typically developing children provided overall statements like, “That was interesting. I wonder if those animals are cursed or something.” Thus, interventionists could help children with autism by spending some time at the end of each lesson or book to discuss what it means. Children with autism should be challenged with questions like, “So, what do you think was the author’s point in this story?” “Were there any morals to this story?” “What did you take away from this story?” In this way, children with autism could be taught to form integrative connections as they read or think about new concepts in school.

**Implications for Research**

*Methodological Research Implications*

*Sample Considerations.* Future researchers might consider recruiting a larger sample of participants from all three groups: Autism, ASD, and Typical. The groups should be matched for IQ, Gender, and Language ability. This would account for potential spurious relationships more adequately. Further, the children should be asked about their images before and after the assessment. These images could then be scored on the Willard Imagery Observation Scales.
Language Measures. In addition to sample considerations, this study invites future researchers to consider the relationship between the WIOS scores and a standardized language assessment such as the Clinical Evaluation of Language Fundamentals (CELF: Semel & Wiig, 2003). Even though some of the participants in the original study were assessed with standardized language measures, there were simply not enough cases that included the same language test; so this issue was not assessed. However, it would be reasonable to expect that some of the WIOS scores would correlate with standardized tests of comprehension based on the theory of integrative-comprehension imagery presented here (Figure 3). However, future researchers should be aware that other studies have found a weaker-than-expected relationship between imagery scores and standardized comprehension measures (Sadoski, 1983). It is likely that several different language measures, especially those with items that are less verbally-loaded, would have to be considered in this analysis.

Increased Imagery Domains. The Willard Imagery Observation Scales might also be expanded in future studies, which would enhance the reliability of the measure and provide a more comprehensive assessment. In this study, it was found that there were potentially six factors underlying the construct of imagery based on the original 15 items in the assessment. Previous studies have found 2–4 factors underlying imagery assessment (Cain, 2009; Oakhill & Patel, 1991; Sadoski, 1983). If a researcher were interested in making this assessment more comprehensive, it would be important to look at: inference-making (Briton & Gulgoz, 1991; Kintsch, 1994), integrating with background knowledge (Kintsch, 1988), and comprehension monitoring (Gambrell &
Bales, 1986) (See Figure 3 for the theoretical model). All of these skills are known to be important to the study of imagery but were unable to be addressed here because of the limited item set available in these archived data.

**Details Domain.** A challenge of this study that might be addressed by future researchers is the Details domain. Initially, a Details domain was included that held the items: sensory, setting, size or shape, time of day, number, and positioning/placement. However, the details domain was highly unreliable during the pilot tests and the entire domain was deleted from the scale. It was determined that the details measured here may be too unrelated to hang together as a construct. However, this is still a potentially important domain to assess based on the theory of imagery (Sadoski, 1983) and autism (Nuske, 2010). It may be possible to include the Details Domain as a single item, rather than a construct. It may be that future research could expand the Details domain into more comprehensive sub-domains. For example, the setting item could be expanded to include the physical setting, the time of day, the month and year of the story, and the season. In this way, it is possible that each potential ‘detail’ of the story could be captured by a common construct, which may enhance the reliability of this domain. Future researchers might consider how to better measure this important; yet somewhat elusive, area of imagery.

**Sensory Item.** The sensory item should surely be considered as an opportunity for further research. Although, the sensory item showed adequate reliability, it was deleted from the scale with the Details domain (as it was the only item loading consistently on the factor). However, this item is indeed important to the study of imagery. Bell &
Binetto (2006), recognized imagery experts believe that one of the primary aims of imagery instruction is to address the sensory area; claiming,

…the [goal] is to bring the sensory information of imagery to a conscious level so it can then be accessed as a sensory tool and integrated with language to establish dual coding. (p.5)

Thus, sensory concepts such as the sights, smells, and sounds provided in a story should be captured in an imagery assessment. Future researchers might consider how to better capture this construct.

Revise items. The inconclusive results of the actions and emotions items, provides a potential direction for future research. The items tended to show variable reliability and to cross-load on the factors. Secondly, none of the predictor variables in the model were significant (IQ, Age, Gender, Diagnosis). It could be that Diagnosis is not a significant predictor on the actions item because children with autism do tend to provide factual information (such as events) as well as typically developing children; which would be consistent with past research (Nuske & Baven, 2010). The variables of Age, Gender, and IQ were also non-significant predictors of scores on the actions item; only Gender was significant on the emotions item. Thus, it is possible that either the items are not well defined or are related to a factor that was not included in the model. Future researchers could consider other ways of measuring ‘actions’ and ‘emotions’ more effectively.

A New Story. Another opportunity for future research is to try this same assessment out with different books. This study only used a book called Tuesday (Wiesner, 1991) because this happened to be the book that was used in the longitudinal
study from whence these data were collected. Research shows that other books like *Little Red Riding Hood* and *Three Billy Goats Gruff* include more dynamic characters and vivid imagery. Strickland and colleagues suggest that stories like these would allow the children to: share feelings, focus on the motives of the characters, and understand the sequence of events. Thus, a book of this nature might have been more effective (Strickland et al., 1985). Therefore, the use of a new story provides an opportunity for future research.

**Theoretical Research Considerations**

*Play Experiences.* Although not assessed in this study, it might be interesting to look at some of the childhood experiences that could have an impact on imagery. For example, children who engage in more symbolic and pretend play, may also show better imagery on a measure like the WIOS. Rutherford et al. (2007) demonstrated that pretend play is impaired in autism. It may be that children who engage in fewer pretend play scenarios have limited practice visualizing, and may show a resultant impairment in imagery skills. Thus, it would be interesting to assess a child’s level of pretend play and compare these results to an imagery scale like the WIOS.

*Problem Solving.* Similarly, children who do not initiate imaginary play with peers or who show a lack of creative play might also demonstrate limited problem solving ability. Children with good social skills are able to solve problems and resolve disputes during play. Children with autism tend to have limited social skills; and generally, limited accesses to these play opportunities (Rutherford et al., 2007). Although early research has shown a relationship between problem solving and imagery (Shaver, Pierson, & Lang,
1975), there was not enough data in this study to evaluate the relationship effectively. It may be that children with autism, who also show impaired problem solving ability, would achieve lower scores on the WIOS. It also may be possible to increase problem solving ability through direct instruction; and this increased understanding may have an impact on imagery. Future researchers might evaluate the relationship between problem solving and imagery.

Executive Functioning. Children with autism tend to have impaired executive functions (Rutherford et al., 2007). Executive functions include such skills as: problem solving, sequencing events, planning, monitoring progress toward goals, metacognition, shifting between activities, and inhibiting emotional responses. The role of executive functioning was evaluated, to a degree, during the pilot study. Indeed, children with autism tended to show significantly poorer executive functions in this study; especially in the area of metacognition, compared to neurotypical controls (as measured by the Behavioral Rating Inventory of Executive Function: BRIEF). This weakness in metacognition has been uncovered in previous research (O’Connor & Klein, 2004). It may be that children with weaker executive functioning show lower scores on the WIOS. Research indicates that children with autism have difficulty sequencing because of impaired executive functions (Rutherford et al., 2007). However, executive functioning data were not available on enough of the participants to include it as a covariate. Future researchers might consider the relationship between executive functioning and imagery.

Screen Time. Many children today spend less time playing and an undue amount of time watching TV. By the time children today graduate from high school, they will
have spent more time watching TV than in the classroom (American Academy of Child and Adolescent Psychiatry: AACAP, 2011). One unfortunate consequence of this screen time is that children who watch TV tend to read fewer books (AACAP, 2011). Children generally practice imagery skills when reading or communicating verbally. However, these skills are less likely to be practiced during passive screen time. Perhaps, children who spend a great deal of time watching TV, may have limited imagery because these cognitive abilities have not been practiced routinely. Thus, the ‘screen’ may take place of the child’s imaging centers, which could impede a child’s own ability to visualize. In contrast, children who are read to frequently and do not have as much screen time, may show increased imagery abilities. Future researchers might explore the relationship between screen time and imagery.

**Imagery Interventions.** Another significant opportunity for future research includes development of a new imagery intervention program. While the scale might be improved with more items in future research, the WIOS scale already pinpoints several areas for intervention. For example, an imagery intervention for autism should include direct teaching in all of the areas identified as weaknesses in this study: characters, actions, sequence, gestures, cause-and-effect, perspective taking, and integrative statements. If indeed an intervention were developed to specifically address these deficit areas, it is possible that children with autism stand to benefit substantially. Children with autism might learn to communicate more effectively; show improved social skills, and demonstrate increased academic achievement. Thus, the development of an imagery intervention for children with autism is a promising direction for future research.
Limitations

There are a number of limitations in this study. First, there were problems with using an archival data set. The researcher did not have access to the participants. As such, it was not possible to ask the children about their images. It would be important to know whether or not they could describe their own images apart from the pictures in the story. It would have been helpful to ask the children about specific story elements to gather a more accurate assessment. For example, a researcher could ask the child at the conclusion of the story, “What happened first, next, and last in your story?” Additionally, the researcher could ask, “What was the main idea of the story?” or “Taken together, what basically happened?” Although this type of analysis was not possible in this study, it would be important to consider in an imagery assessment.

Some studies have begun looking at imagery this way (Gambrell & Bales, 1986; Joffe et al., 2007; Sadoski, 1983). Researchers have simply instructed participants to ‘make pictures in their heads; ‘read the stories to them, and then asked them what they pictured at the end (Sadoski, 1983). These studies found that children with poor comprehension told stories that were accurate, in that they included all of the correct story events, but they were not well integrated or correctly sequenced (Nuske & Baven, 2010). Children with autism were tested for comprehension using a story –retell task; finding, that they recalled stories that were disjointed, rather than sequenced into a meaningful pattern (Nuske & Baven, 2010). While this research is promising, it could be extended by using a more standardized assessment like the WIOS paired with a detailed
imagery interview. This would allow researchers to understand which part of the images are impaired in autism and could guide intervention.

A further limitation of this study is the limited number of items and domains that were included in the final version. The instrument only included a small number of items (8 items) and domains (2 domains). Due to the constraints of using an archival data set, several items and domains could not be captured (i.e. inference-making, linking to background knowledge, comprehension monitoring). Further, the Details domain was deleted because of poor reliability. It would be helpful to consider adding items to the details domain that are more closely related to each other; whereas, the WIOS detail items were more randomly distributed. There also may be a wide variety of other items that could be included in the instrument if the participants were interviewed or assessed in person, rather than by way of observation of the video-tapes. Future researchers might explore the addition of an interview protocol; the inclusion of more items, and the capturing of more imagery domains.

The lack of a standardized language measure as a covariate was a limitation as well. The reason this was not included is that a significant portion of the sample was not administered the same language test. In the original study from whence these data were harvested, some of the children were assessed with the Clinical Evaluation of Language Fundamentals, Form 1; some were assessed with Form 2; some were assessed with the Peabody Picture Vocabulary Test, and some were assessed with the Child Communication Checklist. Further, the participants were not assessed using one of the subtests of the CELF that assesses story comprehension.
The tests that were given were appropriate based on the various language abilities and IQ levels of children in the original sample. However, the lack of these data shows up as a limitation in this study. It was not possible to tease out the effects of language ability as distinct from overall cognitive ability (IQ). Future researchers could use the WIOS in conjunction with a standardized comprehension test to see whether or not the scores correlate significantly.

Further, the fact that language ability is a part of the diagnosis of autism but not of Asperger’s introduces some spurious relationships into the model. That is, it may be that differences between the performance of children with autism compared to those with Asperger’s may be due to the idea that children with Asperger’s generally have better language, thus expressing their images better than those with autism.

Another limitation was uncovered in the study of gesture use. In this sample, the majority of children were of American, White, Non-Hispanic descent. Thus, the ‘gesture’ item on this scale was developed based on American common culture. For example, head-nods mean ‘yes;’ gesturing the hands in a circle means, ‘round.’ However, this gesture item may not be appropriate in other cultures where gesture use is less commonplace or where gestures are ascribed different meaning. It is possible that the item could be rewritten if the WIOS is used in different populations of examinees. It is also may be that the item could be revised such that it is more culturally sensitive.

Finally, sample size (N=71) may have limited the findings to a degree. The researcher was not able to use Structural Equation Modeling or other large-sample methods due to sample size. Further, the sample size of the typical group was particularly
small (N=23). As such, a case in the typical group with an especially high IQ, for example, may inflate the effect of IQ in the regression analysis. This small typical sample also had a much higher proportion of girls as compared to the autism sample. This could also have impacted results; thus, overestimating the effect of gender. It should be stated that a clinical sample like this one is rarely achieved in other studies. The high number of clinically diagnosed children on the autism spectrum (N=48) is a significant strength of this research. However, more significant results may have been uncovered in a larger, more evenly distributed sample.

**Contributions**

Even in light of the aforementioned limitations, this study contributes meaningfully to the science of imagery in autism. First, there is a contribution to the understanding of imagery in the autism population. In almost every regression analysis, the IQ variable was rendered insignificant with the addition of the Diagnosis variable. The results of this study indicate that children with autism might indeed uniquely struggle with integrative-comprehension imagery.

There were interesting findings in this study regarding scores on the *Main Idea Domain*. This research is consistent with the Weak Central Coherence Account, which provides that children with autism tend to miss out on the ‘gist’ of a story (Diehl et al., 2006; Nuske & Baven, 2010). Results of this study are also consistent with research regarding children with poor imagery. Imagery research indicates that children with weaker imagery skills struggle with part-to-whole relationships (Bell, 1991b; Bell & Ortiz, 2006). Bell and Ortiz explain about children with impaired imagery that,
They have a tendency to process parts, more than, or rather than wholes. They get details rather than the big picture (p.10).

Findings from this study are consistent with this theory. These consistent results are an important contribution to the science of imagery in autism because it may be that children with autism could be taught to identify the main idea when they read and then to incorporate that main idea into their own narratives. This type of intervention could allow them to improve their reading comprehension and communication.

The other important contribution was the significance of Diagnosis in predicting scores on the Integration Domain. This study is consistent with the ‘functional under-connectivity’ theory of imagery in autism. There is evidence here to support previous research suggesting a lesser degree of activation between the areas of the brain that process images and those that process language in the brain of autism. This finding offers a significant contribution to the science of imagery in autism. If in fact their brains are lacking connectivity between these structures, it may be possible to forge new connections through rehearsal and explicit teaching. Thus, if interventionists were direct and clear about how the images in the story connect with the words, children with autism may develop the ability to integrate their narratives effectively. This could improve their social communication, reading comprehension, and academic achievement.

A final contribution of this research is with regard to the scores of children with autism on the following items: characters, sequence, gestures, causal connections, perspective taking, and integrative statements. Children with autism told narratives that lacked important elements such as the characters and the sequence of events. Children
with autism told stories without identifying how the characters were thinking or feeling. Finally, their stories lacked causal connections and integration. It is possible that some of the speech and language difficulties experienced by this population are rooted in imagery. Thus, an imagery assessment like the WIOS can be used to determine whether or not imagery interventions are indicated.

Scientists and practitioners should consider this research when attempting to provide for the needs of children with integrative-comprehension problems. The WIOS assessment could be used for typically developing children with a suspected weakness in imagery. Research shows that imagery problems have a significant impact on comprehension. Children with poor comprehension show reduced academic performance. Thus, results of a quick imagery assessment, such as the WIOS, might indicate appropriate instructional and educational interventions for children in the general education population.

The literature indicates that imagery weaknesses are amenable to instructional support and intervention; thus, opening up a variety of opportunities to enhance and inform educational practice. Children, who struggle significantly with comprehension tasks, should be assessed for imagery skills. If indeed these challenges are found, interventions should be evaluated and employed, from which children with comprehension deficits stand to greatly benefit. The findings herein, thus, provide substantial opportunity to enact meaningful change for children with and without autism.
References


159


161


Appendix: WIOS Manual

WIOS MANUAL

The purpose of this instrument is to understand how well a child can describe images and stories. Foundational to imagery for integrative-comprehension is the ability to describe an image on the printed page. Teachers who instruct children to use imagery for comprehension often ask children to describe printed pictures as the first step in the learning process. As the examiner, it is important to keep one goal in mind. The child’s descriptions should be clear enough that the image in the examiner’s mind matches that of the child’s, such that it would not be necessary to read the book to get a thorough portrait of the images in the story.

When coding the child’s responses, it is acceptable to code one response in two places. For example, if the child says, “he’s mad because he can’t fly;” this response would score points on both the Emotions and the Causal Connections items. Generally speaking, in addition to the guidelines below, if a child gives a response that is approaching the goal of the item or approximating an answer, that response should receive 1 point. That item should remain scored as 1 point until the child gives a fully clear and coherent 2 point response. For example, if the child was asked why the man feels scared and the child responded, ‘cuz there’s frogs;’ this is a 1 point response under causal connections because the connection is loose and unclear. However if the child later said, “the frogs are afraid because the dogs are chasing them;” the score becomes a 2 point.

MAIN IDEA DOMAIN

The main idea domain is a measure of a child’s ability to provide narrative descriptions centered around the essential concepts in the story. The Main Idea domain includes characters, actions, and sequencing. The purpose for this domain is to see whether or not a child can give all of the most important information about a story, such that the listener has a basic idea of what happened.

1. CHARACTERS

| 1 pt: Two animals+ | Ex. "dogs" "pigs" "birds" |
| 2 pt: Animals and human character | Ex. "grandma" "the guy" "a man" "he" + animals |

Characters are the central figures in the story. The characters must be live, active agents, rather than objects. The child receives 1 point for mentioning at least two animals. He or she receives two points for mentioning two animals and a human character.
SAMPLE RESPONSES:

1 pt: 2 or more of these: dogs, pigs, birds, horse, turtle, cat
2 pt: 2 or more animals + a human character: a person, grandma, grandmother, old lady, someone, somebody, girl sleeping, all the people, enemies, the guy

2. ACTIONS

| 1 pt | 2 or more actions or motions | Ex. | "running," "chasing," "falling"
|------|-----------------------------|-----|----------------------------------|
| 2 pt | action + object acted upon   | Ex. | "dogs were chasing the birds"
| 2 pt | Ex. "The man was eating a sandwich" "The pigs were messing up the laundry"

This item is intended to measure the child’s ability to describe what the characters are physically doing in the story. 1 pt responses must include more than one action. 2 pt responses must include the action plus the object or character being acted upon.

1 pt: 2 or more of these: “running”, “chasing”, “falling”, “flying”, “crashing”, “sleeping”
2 pt: “the man was eating a sandwich;” “the frogs were messing up the laundry;” “the dogs were chasing the birds” “the birds were sitting on the telephone wire;” “the frog is changing the channel on the TV”

3. SEQUENCE - TEMPORAL ORDER

<table>
<thead>
<tr>
<th>1 pt</th>
<th>uses 1 connecting word: next, then, after, and</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 pt</td>
<td>Two connecting words, then ______ then ______</td>
</tr>
</tbody>
</table>
| 2 pt | Ex. "They run up the tree and then hop home and then swim again"

The sequence / temporal order item is a measure of the child’s ability to connect actions into a logical string of events. For example, one child said, “The frogs get sheets and tie them around as capes then fly to another house.” This would be a two point response because it incorporates two connecting words.

1 pt: “Then they are flying all over;” “Frogs are flying by the window while the man is eating his sandwich;” “one frog got covered and then said who turned out the lights;” “Then they went to another house;” “they go back to the pond;” “and then pigs did it.”

2 pt: “Then, they are flying and then this frog is hanging off the shingles;” “and then they are pulling on stuff and the grandma is sleeping” “now they are super frogs and now they are all watching TV”

INTEGRATION

The Integration Domain measures the general skill of thinking about how story events connect and relate to each other. Rather than simply relaying facts or providing details, children who can integrate story events demonstrate that they understand how to integrate information from various parts of a story into their own narratives. This domain includes:
Emotions, Causal Connections, Perspective Taking, and Gestures. The emotions item requires children to provide a rationale for how the person feels, which is why emotions falls under integration. The Gestures item was selected because literature shows that children who gesture are attempting to communicate images. Further, gestures are often used to show how events are related and connected. Causal Connections and Perspective Taking are included because these items require examinees to integrate across the story.

4. EMOTIONS

1 pt: provides 2 or more feeling words  
Ex. "angry" "mad" "happy"

2 pt: Feeling with rationale  
Ex. "he is scared because the dogs are chasing him"

The emotions item captures the feelings and mood of the characters in the story. The mood should not be the child’s mood in reading it, such as ‘this is freaky.’ For 1 point the child needs to provide two or more feeling words. For a 2 point response, the child must provide a feeling with a rationale for why the person feels that way. The rationale should be clear and sensible. Award 0 pts for any response that is a direct repeat of the examiner’s words. For example, if the examiner says, “I think he feels scared” and the child says, “Yeah he’s scared;” that is a 0 pt answer.

1 pt: 2 or more of these: angry, mad, happy, scared, surprised, startled, looks mean

2 pt: “He’s mad because he can’t fly” “He’s mad because he’s been up all night” “he’s mad because he has to wait again” “he’s surprised because he doesn’t know frogs can fly” “they are scared because they are being chased by the dog”

5. GESTURES

1 pt: simple gesture  
Ex. head nod, point, shrug

2 pt: descriptive gesture, use body to describe  
Ex. gestures ‘putting on capes’

The gestures domain is built on the literature-based theory that children with stronger imagery will be more apt to describe their own images with gestures. This item is for gestures directed toward the examiner. 1 pt responses include simple gestures such as head nods and shoulder shrugs. 2 pt responses are for gestures that are used to describe the story to the examiner such as pulling ones hands over the shoulders to demonstrate the frogs putting on capes.

1 pt: head not, point, shrug

2 pt: gestures hands in large circle to describe ‘exploding;’ gestures wrapping of sheet around waste to show ‘putting on a sheet like a diaper;’ gestures hands over shoulders to demonstrate ‘putting on capes;’ gestures a person’s teeth chattering or biting one’s nails to show that the person in the story feels nervous or scared.

6. CAUSAL CONNECTIONS

1 pt: Loose causal connection of events or ideas  
Ex.: "its cuz he’s scared"

2 pt: Clear causal connection  
Ex.: "They are scared because dogs are chasing them"
Causal connections is an item measuring the child's ability to give a clear rationale for a character's feelings or actions. If the examiner asks why something is happening and the child gives a response including the word 'because,' this receives 1 pt if the causal connection is loose or unclear. For example, if the examiner asks, why is the man looking out the window and the child responds, “cuz he's scared;” this receives 1 pt. If the response includes a clear causal connection, it receives 2 points.

1 pt: “it's cuz he's scared” “cuz he is wondering” “cuz they want to have fun” “because they are getting chased” “fall off their lily pads because it's daytime”

2 pt: “the frogs had to stop flying because they can't see;” “that frog is grumpy because he wants to fly and he can't” “they lost their lily pads so they can't fly now” “I know it's a dog because of its leash” “dogs were gonna eat the frogs so they ran off as fast as they can”

7. PERSPECTIVE TAKING

1 pt: Takes character's visual point of view Ex: "He sees frogs through the window"
2 pt: Takes character's mental state Ex.: "He's wondering why lily pads are everywhere"

Perspective taking is an item measuring a child's capacity to take the character's point of view. Any response that reports the child's perspective receives 0 points. For example, if the child says, “I am looking at it from above the house;” this is a 0 pt response. The child receives 1 point for taking the character's physical point of view and 2 points for taking on the character's mental state.

1 pt: "he sees frogs through the window" "the turtle saw a frog;" "the wolf saw a lily pad;" "someone saw the frogs" "they all see lily pads"

2 pt: “He’s wondering why lily pads are everywhere;” “And the guy says...mmmmm...what are all these lily pads doing here?"; “the man didn’t know the frogs could fly;” “And the man said, what is this mess down here?”; “she doesn’t notice the frogs because she is asleep;” “he is thinking, why are they flying?” “Someone didn’t want the frogs in the house” “He’s thinking what in God’s name are these frogs doing here?” They are saying, “Someone help, it’s flying pigs;”

8. Integrative Comments

1 pt: only one comment at the end
2 pt: 'I wonder what happens next', 'It would be neat to fly,' 'That was a funny book'

The integrative comments item is an assessment of the child's ability to put different elements of the story together. For example, at the end of the book when the pigs fly, some of the children said, “and it starts all over again with pigs.” A statement like
this indicates that the child is integrating the story events. If the child provides one integrative comment at the end, he or she receives 1 pt. If the child provides two or more integrative comments, he or she receives 2 points. If the child provides a loose integrative statement, score a 1 point. For example, “I don’t like books” or “That was silly” or “that’s cool” are 1 point answers.

**1 pt:** “The end;” “They lived happily ever after;” “that’s crazy”

**2 pt:** Two or more of these types of statements: "I bet the frogs are still back in the pond;" "I think I know what will be next. Maybe squirrels;" "I don’t think pigs could fly because they don’t have anything to float on;" "It would be neat to be able to fly;” “I think they are cursed or something;” “It’s like aliens are taking over” “There were flying frogs and now there are flying pigs” “I think the frogs will strike again” “It’s weird that they can fly on lily pads”