THE INITIAL DEVELOPMENT AND CONTENT VALIDITY OF AN ASPERGER'S SYNDROME SELF-SCREENING INSTRUMENT FOR ADULTS

A Thesis Submitted to the College of

Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the degree of Master of Education in the

Department of Educational Psychology and Special Education

University of Saskatchewan

Saskatoon, Saskatchewan

By

Karen Fung

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of Department of Educational Psychology and Special Education
University of Saskatchewan
28 Campus Drive
Saskatoon, Saskatchewan
S7N 0X1

FOREWORD

Results of this study will assist in the development of a future Asperger's Syndrome self-screening instrument for adults. I would like to thank Dr. Janine Montgomery and her research team at the University of Manitoba for their willing participation

DEDICATION

This thesis is dedicated to all the people who have supported me all the way through. Thank you Mom and Dad for always being supportive no matter what I do and what my decision is. You always give me the freedom to try and explore anything that interests me. Thank you for all the emotional and financial support you have provided throughout the years. You have taught me to become the strong, independent high-achiever I am today. Thank you for all your unconditional love. You have given up so much to provide a better life for my sister and I.

Thank you to my long-term boyfriend, Chris for all your support throughout all the years. Thanks for cooking for me when I was stressed and tired. Thank you for understanding my desires to achieve my goals and went off to Saskatoon myself leaving you alone in Toronto. Thank you for all the motivations you gave me to work hard for the future, and hopefully, our future.

Thank you to all my friends who care so much about me. I feel loved when I am around with you guys! To my best friend Winnie who went through the same process with me. Thank you for listening when I need someone to talk to, you always know me so well.

In loving memory of my grandpa, and Aunt Margaret who passed away when this thesis was still in progress. You will always be missed.

ABSTRACT

This research addresses the lack of an existing psychometrically sound Asperger's Syndrome self-screening instrument for adults. Initial instrument development procedures were carried out by creating an item pool using existing Asperger measures. Items were rewritten following common item writing rules suggested by several researchers as reference. Five new domains were created and the items were re-categorized. Expert panel of nine judges specialized in Psychometrics, ASD, School Psychology, and Speech-Language Pathology were asked to rate the relevancy of items to their domains in order to obtain evidence of content validity. These experts were chosen because of their relatedness to Asperger's Syndrome and their expertise in instrument development. First, the quality of the judges' ratings were examined to identify any aberrant judges. Ratings were then analyzed using the remaining six judges using three descriptive and three quantitative methods to examine the representativeness and relevancy of each item to their domain. A total of 55 items were identified as satisfactory by the judges. The second part of this study was to compare the content validity analytical methods. It was concluded that the percentage agreement, the content validity index (CVI), and the content validity coefficients (VI_K) were the best methods to use in selecting the satisfactory items. This research aims to bring more attention to the importance of psychometric properties in measures for the Autism Spectrum Disorder field. It also hope to shed some light on which content validity analyses would best be used under certain circumstances. Limitations of study and future directions were also discussed.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the all the help and guidance of my supervisor Dr. Laurie Hellsten. You are truly an amazing supervisor that I cannot ask for a better one. Thank you for being extremely patient with me (and my writing) by spending numerous of hours editing this thesis with a very tight deadline. I really appreciate your support and all the opportunities you have provided me. You helped me with directions and suggestions when I was lost, and you prepared me very well to enter the Psychometric field. I am really grateful to have a supervisor like you.

Thank you Dr. Laureen McIntyre for being my committee member. You helped me a lot in editing this thesis, revising the items, and categorizing the items. You provided me with lots of suggestions with your expertise in Speech-Language Pathology, Special Education and even APA style.

Thank you Dr. Todd Rogers for allowing me to bug you even when my thesis is not related to course work. Your offer to help was truly appreciated.

And lastly, thank you Dr. Janine Montgomery for providing me this opportunity to work with you. It is my pleasure to be part of your project.

TABLE OF CONTENTS

CHAPTER 1 ..... **1**

1. INTRODUCTION 1
1.1 Research Background *1*
1.2 Purposes of Study *6*
1.3 Significance of Study *7*
1.4 Definitions *7*

CHAPTER 2 ..... **9**

2. REVIEW OF THE LITERATURE 9
2.1 Asperger's Syndrome *9*
2.1.1 What is Asperger's Syndrome? 9
2.1.2 Prevalence of AS 11
2.1.3 AS Diagnostic Criteria 12
2.1.4 Other AS Characteristics 12
2.1.4.1 Comorbidity in AS. *14*
2.1.5 AS and Autism 16
2.1.6 Causes 17
2.1.7 Effects/Impacts of As for Individuals with AS 18
2.1.8 AS Measures 19
2.1.9 Instrument Critique 19
2.1.9.1 RAADS *20*
2.1.9.2 AQ. *23*
2.1.9.3 AAA. *28*
2.1.9.4 The Australian Scale for Asperger Syndrome – Adult Version.. ..... *32*
2.1.9.5 ASDI. *34*
2.1.9.6 GADS 37
2.1.9.7 Overall Quality of the Scales. *40*
2.2 Scale Development *40*
2.2.1 Item Writing 41
2.2.2 Reliability and Validity 42
2.2.2.1 Reliability. *42*
2.2.2.2 Validity. *44*
2.2.2.3 Content validity. *45*
2.2.2.4 Methods of Analyzing Content Validity. *45*
CHAPTER 3 DEVELOPMENT OF ITEM POOL 47
3.1 Item Writing and Categorization 47
3.1.1 Methodology 47
3.1.2 Creation of Domains 49
3.1.3 Domain Definition Results 49
3.1.4 Item Writing and Categorization Results 53
3.2 Preliminary Item Screening and Judgment 53
3.2.1 Methodology 53
3.2.2 Results 54
3.3 Item Content Rating Review Forms 54

3.4 Expert Judgment 55
3.4.1Methodology 55
3.4.1.1 Participants. *55*
3.4.1.2 Expert Judgment Procedures. *57*
3.4.2 Results 57
3.4.2.1 Participants. *57*

CHAPTER 4 59

4. CONTENT VALIDITY – ANALYSIS OF RATINGS 59
4.1 Methodology 59
4.1.1 Identifying Aberrant Judges 60
4.1.1.1 Detecting the lie items. *60*
4.1.1.2 Inter-judge agreement (JDM) *61*
4.2 Results 62
4.2.1 Identifying Aberrant Judges 62
4.2.1.1 Detecting the lie items *62*
4.2.1.2 Inter-judge agreement (JDM). *63*
4.2.1.3 Removal of experts. *67*
4.3 Descriptive Analysis 68
4.3.1 Methodology 68
4.3.2 Results *69*
4.3.2.1 Item ambiguity. *70*
4.3.2.2 Percentage agreement. *71*
4.4 Quantitative Analysis 74
4.4.1 Methodology 74
4.4.1.1 Content validity index *74*
4.4.1.2 Content validity ratio *74*
4.4.1.3 Content validity coefficient *75*
4.4.2 Quantitative Analysis Results 76
4.4.2.1 Content validity index *76*
4.4.2.2 Content validity ratio *74*
4.4.2.3 Content validity coefficient *75*
4.4.2.4 Agreement across all methods *75*
4.4.2.5 Item difficulty *75*
4.4.2.7 Comparison of analytical methods *89*

CHAPTER 5 91

5. DISCUSSIONS AND CONCLUSIONS 91
5.1 Summary *91*
5.1.1 Purpose of Research 91
5.1.2 Procedures 92
5.1.3 Discussion of Findings 92
5.1.4 Limitation of Study 94
5.1.5 Future Directions 95
5.2 Conclusion *97*

REFERENCE 99

APPEDIXES

Appendix A. Diagnostic Instruments 109
Appendix B. 40 Item-Writing Suggestions by Frey et al. (2005) 112
Appendix C. Methods of Content Validity 114
Appendix D. Modification of Items: Original Items and Their Associated Measures 115
Appendix E. Item Content Rating Review Forms 132
Empathy 132
Stereotyped and Restricted Repetitive Patterns of Interests and Activities 134
Deficit in Social Communication Skills: Other Problems in Communication Skills 136
Deficit in Social Communication Skills: Problems in Pragmatic Language 138
Sensorimotor and Stereotyped Patterns of Behaviours 140

LIST OF TABLES

Table 3.1 Distribution of Judges' Expertise....58

Table 4.1 Identification of Lie Items by Judges....63

Table 4.2 Inter-Judge Agreement....64

Table 4.3 Comparison of Methods on Identifying Aberrant Judges....65

Table 4.4 Inter-Judge Agreement with Judge 1 Excluded....67

Table 4.5 Median of Items by Subscale....70

Table 4.6 Item Ambiguity....71

Table 4.7 Percentage Agreement....73

Table 4.8 CVI Results....77

Table 4.9 CVR Results....78

Table 4.10 VI_K for Each Item....79

Table 4.11.1 All Ratings for Empathy....81

Table 4.11.2 All Ratings for Stereotyped and Restricted Repetitive Patterns of Interests and Activities....82

Table 4.11.3 All Ratings for Sensorimotor and Stereotyped Patterns of Behaviours....83

Table 4.11.4 All Ratings for Deficit in Social Communication Skills: Problems in Pragmatic Language....84

Table 4.11.5 All Ratings for Deficit in Social Communication Skills: Other Problems in Communication Skills....85

Table 4.12 Items Remaining after Item Deletion....87

Table 4.13 Categorization of Judges' Comments....89

Table 4.14 Characteristic Summary of Analytical Methods....90

Table A1 Examples f Existing ASD Measures....107

Table A2 Examples and Descriptions of AS/HFA Measures....108

Table A3 Diagnostic Tools Commonly Used.....109

CHAPTER 1

1. Introduction

1.1 Research Background

Within the context of developmental disabilities, typically five disorders are categorized under the umbrella term of Autism Spectrum Disorders (ASD). Autism or Autistic Disorder (AD) is a popular term used by the general population when the idea of ASD is brought up, as it represents the most portrayed case of ASD. However, other labels along the spectrum of ASD do not seem to attract as much attention; these labels include Rett syndrome, Childhood Disintegrative Disorder (CDD), and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS). Rett syndrome is a progressive neurological disorder that only affects girls (Van Acker, 1991). Apparently normal girls at the age of 6 to 18 months with this disorder begin to display "a progressive loss of cognitive and motor skills as well as the development of stereotypic hand movements" (Van Acker, 1991, p. 1) until age of three. These girls also begin to display a disinterest in other people and things, while maintaining eye contact (Holm, 1985; Trevanthan & Naidu, 1988; Witt-Engerstrom, 1987; Van Acker, 1991). Childhood Disintegrative Disorder was actually identified earlier than AD, when named *Infantile Dementia* by an Austrian special educator Theodore Heller in 1908 (Hendry, 2000). Like the girls with Rett syndrome, children with CDD initially showed normal development, until age 3 to 4, when the mood, behaviours and intellectual functioning begin to progressively worsen, and these children later lose their "receptive and expressive language capabilities, developmental incontinence, and eventually require custodial care and treatment" (Hendry, 2000, p. 78). Individuals classified in the PDD-NOS category do not meet all the requirements of AD, but display similar

characteristics of AD (Mayes, Volkmar, Hooks, & Cicchetti, 1993). Classic Autism (or AD) and Asperger Syndrome (AS) are often seen as very similar to one another, except for the language component. In comparing the two, AD is on the more severe side, in which individuals display more intense symptoms in addition to a delay of language development (American Psychiatric Association, 1994; Ozonoff et al., 2003). Ever since Kanner (1943) introduced the term *infantile autism*, numerous studies have been conducted to gain a better understanding of AD. On the other hand, AS did not gain much attention until Lorna Wing (1981) introduced the term to describe cases of individuals portrayed by Hans Asperger in 1944. Thus, AS is considered a *newer* field and the focus of this research will be on AS in an attempt to try to bridge some of the *missing gap*.

Individuals with AS are described as: having impaired social interactions; being incapable of demonstrating non-verbal communication; being resistant to change; often showing repetitive behaviours; having poor motor coordinating skills; and showing intense and limited interest (Wing, 1981). However, they also seem to acquire speech normally (Wing, 1981). After 23 years of introduction, AS finally gained a place in the Fourth Edition of the Diagnostic and Statistical Manual of Mental Disorder (DSM-IV) in 1994 as a separate diagnosis listing under the category of Pervasive Developmental Disorders (Ozonoff, Rogers, & Hendren, 2003). One main area that distinguishes those with AS from others with AD is speech development. In order to receive a diagnosis of AS, the individual must demonstrate normal developing rate of speech, in which single words must be used by age 2, and communicative phrases are shown by the age of 3 (American Psychiatric Association, 2000). Otherwise, the person will be classified as having AD.

Estimations of AS prevalence come in a very wide range. AS prevalence ranges from every 2 per 10,000 people (Fombonne & Tidmarsh, 2003) to every 7.1 individuals per 1000 people (Ehlers & Gillberg, 1993). In Canada, AS is estimated to occur in about five of every 10,000 Canadians (Autism Society Canada, 2005). This number also seems to be increasing as a result of increasing awareness in research and in the public at large (Ozonoff et al., 2003). The large range of prevalence may also be due to different research methodologies and measurement being utilized by different researchers.

In addition to the general characteristics, some researchers examined other differences in AS. Baron-Cohen (2001) suggested that people with autism might show a deficit in Theory of Mind, or the ability to understand others' mind and mental states. For example, these individuals may have difficulty making sense of other people's behaviours (Baron-Cohen, Leslie, & Frith, 1985). In terms of motor coordination, Rinehart, Bradshaw, Brereton, and Tonge (2002) demonstrated that both the group with autism and the group with AS have dysfunction in executive function. Executive function is the capability to "execute mental control necessary for maintaining a problem solving strategy to obtain a future goal" (Lopez, Lincoln, Ozonoff, & Lai, 2005, p. 445). Other characteristics such as differences in sleep patterns (Allik, Larsson, & Smedje, 2006; Liu, Hubbard, Fabes, & Adam, 2006), and attention deficit (Schatz, Weimer, & Tauner, 2002) were also found in children with AS. Comorbidity, or the display of another disorder along with the existence of one is also common in AS; comorbidities may include Obsessive Compulsive Disorder (OCD), Attention Deficit/Hyperactivity Disorder (ADHD), eating disorder, and depression (Gillberg & Billstedt, 2000).

Many theories have been proposed for causes of AS and AD. Causes of AS may be biological (neurological) or may be due to genetics. Genetically, studies have demonstrated that

first-degree relatives tend to show some signs of AS (Ghaziuddin, 2005), and siblings of people with AS people also performed poorly on social cognitive tests (Dorris, Espie, Knott, & Salt, 2004). Neurologically, differences in brain parts such as the frontal lobe, was suggested to be different in people with AS (Stone, Baron-Cohen, & Knight, 1998). Being unable to communicate effectively in society often makes one's life difficult, especially for people with AS. People with AS often have problems such as maintaining jobs and negotiating social relationships in employment, which affects their overall quality of life as adults (Barnhill, 2007; Hurlbutt and Chalmers, 2004). These adults have also been found to have neurological and sensory issues (e.g., oversensitivity to touch and noise), possible problems with the legal system (e.g., crimes), mortality rates, and problems following treatments (Barnhill, 2007).

Diagnosis of AS is mainly made through the use of the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders – Text Revision (DSM-IV-TR; American Psychiatric Association, 2000) and the International Classification of Diseases (ICD-10; World Health Organization, 1993). There is an increasing number of diagnostic and screening measures for AS, and most of them tend to combine AS with High Functioning Autism (HFA) due to a lack of agreement on separation between the two in the field of ASD. For example, Klin, Volkmar, Sparrow, Cicchetti and Rourke (1995) have suggested that the neuropsychological profiles between children of AS and HFA are differed. On the other hand, Manjiviona and Prior (1995) failed to find a significant difference between AS and HFA children using a standardized test on motor impairments. Although there is still a lack of agreement as to whether AS and HFA are interchangeable, examining these AS measures in depth will allow for a review of their psychometric properties. This issue will be further discussed in Chapter 2. The target populations of these instruments are often children and adolescents. Very few instruments have been

developed for adults specifically, and most of these measures are to be completed by practitioners or parents, with no existing self-screening instruments for adults believed to have AS. As AS is a concept still in development from the1980s until this time, adults in the past centuries never had the opportunity to obtain a diagnosis when they were children. It is likely that many of these individuals grew up without proper identification of their long-term communication problems. Therefore, at this present time when experts are becoming more knowledgeable in AS and ASD, a psychometrically sound AS self-screening instrument for adults is essential. This study plays a role in developing this potential instrument. Although self-screen is not a diagnostic tool, it is important to develop such instrument to assist individuals seeking proper diagnosis if needed.

To develop a good instrument is not easy and requires much effort. Following a review of the literature, no rule of thumb or gold standards for instrument development could be obtained other than one developed by DeVellis (2003). DeVellis (2003) suggested an eight-step guideline for scale development. The eight steps involved (1) determining clearly what it is you want to measure; (2) generating an item pool; (3) determining the format for measurement; (4) reviewing the initial item pool by experts; (5) considering inclusion of validation items; (6) administering items to a development sample; (7) evaluate the items; and (8) optimize scale length. The focus of this research study is step 2 through step 4. Factors to consider in the first four steps also include collecting validity and reliability evidence for the scale. Reliability, which often goes along with validity, represents “the proportion of variance attributable to the true score of the latent variable” (DeVellis, 2003, p. 27). It indicates the consistency and stability of a scale. Ways of determining reliability include test-retest reliability, alternative forms of reliability, split-half reliability, internal consistency, and inter-rater reliability (DeVellis, 2003). On the other hand,

validity is an "integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores of other modes of assessment" (Messick, 1991, p. 1). Validity demonstrated how accurate and representative the test content is in covering its' construct. Validity can be combined into a single category of *construct validity*, as suggested by Messick (1991). However, in order to aid instrument development, validity can also be broken down further into four types: content validity, criterion-related validity, construct validity, and consequential validity. Content validity is the main focus of this study. Content validity is "based on professional judgments about the relevance of the test content of a particular domain of interest and about the representativeness with which the set of items deemed to be relevant represents that domain" (Rogers, 2010, p. 231). Evidence of content validity is usually based on subjective and individual expert judgments (Allen & Yen, 1979). There are many ways of assessing content validity evidence in the form of expert ratings, and Hellsten (2008) has suggested that multiple methods should be used to generate a more persuasive conclusion.

1.2 Purposes of Study

To address the problem regarding the need for a psychometrically sound AS adult self-screening instrument, this research study: (1) began to develop an instrument using the first four steps of scale development, by creating an item pool using existing instruments assessing AS; (2) collected content validity evidence using expert judges specializing in Psychometrics, AS, School Psychology and Speech-Language Pathology; (3) utilized multiple methods of analyzing content validity (expert judgments) evidence; and (4) to determine the best method of assessing judgmental results.

1.3 Significance of Study

This research involves the first few steps of instrument development for a future adult AS self-screening instrument. Upon completion of the study, a set of useful items will be created with appropriate content validity evidence. It will hold potential usefulness for the field of ASD in helping adults suspecting of having AS to perform a preliminary screening. The emphasis of psychometric evidence in a scale should demonstrate to the area of ASD that such attention is required. More understanding on the different methods of assessing judgmental analysis will also be gained when comparisons between the methods are made.

1.4 Definitions

The following definitions will be utilized throughout the study.

Autism / Autistic Disorder (AD)

Autism Society Canada (2005) describes individuals with Autism as displaying verbal and nonverbal communication deficits; restricted activities such as following the exact same routine daily; repetitive behaviours such as hand flapping and rocking; cognitive impairment, and deficits in social understanding.

Autism Spectrum Disorder (ASD)

Autism Society Canada (2005) describes individuals with ASD as having particular commonalities in social communication and other characteristics, but their conditions cover a wide spectrum, with individual differences in number and particular kinds of symptoms, the level of severity, the age of onset, the levels of functioning, and their challenges with social interactions.

Asperger's Syndrome (AS)

Ozonoff et al. (2003) describe individuals with AS as possessing social disabilities and

restricted, repetitive behaviours similar to autism, but with well developed language capabilities and normal cognitive functions.

Content Validity

According to Haynes, Richard, and Kubany (1995), content validity refers to the degree that items of a measuring instrument are congruent and reflecting to the construct it is intended to measure.

Pervasive Developmental Disorder (PDD)

According to the DSM-IV-TR (APA, 2000), PDD is a term that umbrellas other disorders characterized by pervasive and severe impairments to communication skills, reciprocal social skills, and/or the display of stereotyped behaviours, interest and activities. All the ASDs are listed under the PDD in the DSM-IV-TR (APA, 2000).

1.5 Organization of Chapters

Chapter Two of this paper will review the literatures of research background, Asperger's Syndrome will be discussed including its history, symptoms, and existing measuring instruments. Then scale development will be discussed emphasizing the importance of reliability and validity, and leads our way to the focus of this study – content validity. Chapter Three focuses on the first part of this research which involves with development of item pool. Methodologies and Results will be described in detail. Then the second part of this study will be demonstrated in Chapter Four on content validity analyses, which also includes the identification of aberrant judges. And lastly, Chapter Five of this paper will ends with discussions and conclusions drawn from the results, limitations of research and future research will also be discussed.

CHAPTER 2

2. Review of the Literature

This chapter examines reviews the relevant related research and discusses characteristics of Asperger's Syndrome such as what it is, its diagnostic criteria, and its difference from AD. This discussion is followed by a critique of six popularly used AS instruments [i.e. Ritvo Autism and Asperger's Diagnostic Scale (RAADS; Ritvo, et al., 2008); the Autism-Spectrum Quotient (AQ; Baron-Cohen et al., 2001); the Adult Asperger Assessment (AAA; Baron-Cohen et al., 2005); the Australian Scale for Asperger Syndrome – Adult version (Meyer, 2000; Attwood, 1998); The Asperger Syndrome (and High Functioning Autism) Diagnostic Interview (ASDI; Gillberg, Gillberg, Wentz, & Rastam, 2001); and the Gilliam Asperger's Diagnostic Scale (GADS; Gilliam, 2001)]. This critique leads to one purpose of this research – the need for a new instrument.

2.1 Asperger's Syndrome

2.1.1 What is Asperger's Syndrome?

Autism Spectrum Disorders (ASD) is classified as a Pervasive Development Disorder (PDD). And under the umbrella of ASD, Asperger's Syndrome (AS) is one of the five disorders along the spectrum. Other disorders along the spectrum include autistic disorder (AD), Rett's disorder, Childhood Disintegrative Disorder (CDD), and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS). AD will be discussed later in this chapter due to its similarity to AS. Rett syndrome is a progressive neurological disorder that only affects girls (Van Acker, 1991). Girls may at first appear to show normal development until the age of 6 to 18 months when they begin to display "a progressive loss of cognitive and motor skills as well as the

development of stereotypic hand movements" (Van Acker, 1991, p. 1) until the age of three. These girls also begin to display a disinterest in other people and things, while eye contact remains intact (Holm, 1985; Trevanthan & Naidu, 1988; Witt-Engerstrom, 1987; Van Acker, 1991). Along the line of the spectrum, CDD was actually identified earlier than AD, named *Infantile Dementia* by the Austrian special educator Theodore Heller in 1908 (Hendry, 2000). Like Rett syndrome, children initially showed normal development, until the age of 3 to 4 when the mood, behaviours and intellectual functioning begin to progressively get worse. These children later lose their "receptive and expressive language capabilities, development incontinence, and eventually required custodial care and treatment" (Hendry, 2000, p. 78). When individuals show symptoms of PDD but do not exactly meet all of the requirements of ASD, then PDD-NOS may be diagnosed (Mayes, Volkmar, Hooks, & Cicchetti, 1993). AS is the focus of this study due to its 'recent' status in the field of ASD, and the lack of attention that has been paid to AS relative to the amount of attention paid to AD.

The first few case descriptions of AS were recorded by an Austrian pediatrician Hans Asperger (1944) when he referred to these individuals as *autistic psychopathy* (Ozonoff, Rogers, & Hendren, 2003; Tryon, Mayes, Rhodes, & Waldo, 2006). Later, Lorna Wing (1981) introduced a more neutral term of *Asperger Syndrome* due to the tendency of people equating the term psychopathy with negative, sociopathic behaviours. However, AS did not gain much attention from researchers, and it was not listed as a disorder in the Diagnostic and Statistical Manual of Mental Disorder until the fourth edition (DSM-IV; APA, 1994) came out in 1994 (Ozonoff et al., 2006). Although AS seems to be a modern diagnosis, AS might have existed earlier in history, perhaps as early as the mid-1800s, almost a century before Asperger's period. Koegel (2008) provided some evidence for this by analyzing a short story *Bartleby* written in

1853 by Herman Melville. If current diagnostic criteria for AS were used to assess the behaviours of the main character Bartleby, he would be a perfect example of AS (Koegel, 2008). Bartleby displayed an inability to use nonverbal cues such as maintaining eye contact during social interactions. Emotional and social reciprocity were often absent in Bartleby, and developing peer relationships was also difficult for him. In addition, it was hard for him to share interest, enjoyment or even achievement with other people (Koegel, 2008). All the characteristics illustrated fit well with the problems displayed by an individual with AS. Even if Bartleby was not a true story, Koegel (2008) argued it was possible for Melville to write stories based on what he had encountered. Therefore, AS may have existed for a very long period of time without anyone paying close attention to the issue.

2.1.2 Prevalence of AS

Wing (1981) reported that AS is more common in boys that in girls. Ehlers and Gillberg (1993) found that the male to female ratio of AS was 4:1. Different researchers have demonstrated different prevalence rates for AS across genders. The prevalence of AS ranges from 2 per 10,000 (Fombonne & Tidmarsh, 2003) to 7.1 per 1000 (Ehlers & Gillberg, 1993). This large range may be due to differences in the research methods used by researchers, or the different diagnostic criteria utilized. In Canada, the prevalence rate for AS is about five in every 10,000 Canadians (Autism Society Canada, 2005). The number of AS cases are also increasing, which may be due to an increased awareness and greater responsibility in practitioners for screening, diagnostic and evaluation (Ozonoff et al., 2006). The increase could also be due to a rise in ASD research and the attempt to obtain greater categorizations within each diagnostic label.

2.1.3 AS Diagnostic Criteria

AS is characterized by mild to severe deficits in social interaction and understanding. The individuals affected also tend to have restricted and repetitive behaviours, activities, and interests, yet, their language and cognitive development is not delayed (Autism Society Canada, 2005; Lopata, Thomeer, Volker, & Nida, 2006, Ozonoff et al., 2003). These individuals may display behaviours such as avoidance of eye contact, stereotyped and repetitive motor mannerisms, rigid adherence to nonfunctional routines, persistent preoccupation with parts of objects, and they may also fail to develop normal peer relations, be unable to share enjoyment, and be incapable of showing social or emotional reciprocity (Lopata et al., 2006). The DSM-IV-TR (APA, 2000) diagnostic criteria for AS include: (a) Qualitative impairment in social interaction; (b) Restricted repetitive and stereotyped patterns of behaviours, interests, and activities; (c) The disturbance causes clinically significant impairment in social, occupational, or other important areas of functioning; (d) There is no clinically significant general delay in language (e.g., single words used by age 2 years, communicative phrases by age 3 years); (e) There is no clinically significant delay in cognitive development or in the development of age-appropriate self-help skills, adaptive behaviour (other than in social interaction), and curiosity about the environment in childhood; and (f) Criteria are not met for another specific Pervasive Developmental Disorder or Schizophrenia (APA, 2000, p. 84).

2.1.4 Other AS Characteristics

It has also been suggested that people with AS have a deficit in ‘Theory of Mind’ (ToM). ToM is defined as “the cognitive mechanism that allows an individual to infer the mental states of other individuals in relation to their own and is therefore likely to be associated with perspective taking and the capacity to empathise” (Murphy, 2006, p. 99). Having a ToM deficit

indicates an inability to understand others' emotions and put oneself 'into other's shoes'. Individuals with AS also appear to perceive human relationships differently as well. A qualitative study conducted by Carrington, Templeton, and Papinczak (2003) examined the perception of friendship in AS teenagers via interviews. The participants seemed to have trouble understanding and using language to describe issues in friendship. The participants also showed a "lack of insight into what constitutes friendship" and an inability to understand the reciprocal nature of being friends (Carrington et al., 2003, p. 216). To explore individuals with AS's understanding of complicated emotions, Shamay-Tsoory (2008) examined the ability of individuals with AS and HFA to understand envy and gloating using computerized ToM tasks and emotion tasks. These emotion tasks often required understanding of more than one emotion (e.g., seeing someone happy leads one to feel angry or sad), and it was predicted that individuals with AS and HFA have an impairment with respect to these emotions (Shamay-Tsoory, 2008). Results indicated the subjects had no problem with basic ToM conditions, however, they showed a deficit in recognizing envy and gloating (Shamay-Tsoory, 2008). Another interesting study examined human figure drawing by children with AS. Lim and Slaughter (2008) suggested that the drawings of human figures represented the overall lack of social world interest in these children, and their lower scores on these figures compared to a normal control group were simply a lack of practice in drawing humans. The researchers also found a positive correlation between these drawing scores and a communication sub-scores on behaviour scales (i.e. Vineland Adaptive Behaviour Scales) in the AS but not the control group.

Apart from affective differences in individuals with AS, Schatz, Weimer, and Tauner (2002) have demonstrated that children and young adults with AS may exhibit attention deficit. In comparison to a control group, they found a greater variability in the results of eight

participants with AS on a Test of Variables of Attention (T.O.V.A.) continuous performance test, to which the authors attributed to attention differences. The T.O.V.A is a standardized computer test on attention, in which participants were asked to press a button whenever a target stimulus is seen on the computer screen (Schatz, Weimer, & Tauner, 2002). Sleep patterns also seemed to differ in individuals with AS. Allik, Larsson, and Smedje (2006) found that, in comparison to a control group, parents of children with AS usually report that their children have more difficulty falling asleep, lower sleep efficiency, and lower sleep quality. In addition, these children also spent a longer time awake in bed before falling asleep. However, sleep patterns between the two groups did not differ. On the other hand, Liu, Hubbard, Fabes, and Adam (2006) examined factors that may be attributable to sleeping disturbances in children with AD. Results of the Liu et al. (2006) study showed that the children with AD tend to have either dyssomnias or sleeping disorders that individuals may experience problems with falling asleep or staying asleep, or problems with sleeping excessively. They also tended to display parasomnias or disorders in which sleep would get disrupted (Liu et al., 2006). In addition, younger age, hypersensitivity, co-sleeping, epilepsy, attention-deficit/hyperactivity disorder (ADHD), asthma, bedtime ritual, medication use, and family history of sleeping problems were factors that were found to relate to sleeping problems in children with AS (Liu et al., 2006). It seems the comorbidity of other problems indirectly affect children's sleep.

2.1.4.1 Comorbidity in AS. It is not uncommon for comorbidity to occur in individuals with AS. Gillberg and Billstedt (2000) defined comorbid to occur when one of the following criteria is met: that the two conditions are (a) coincidental; (b) casually directly related, with one condition leading to the other; or (c) the conditions are casually indirectly related, in which other problems are leading to both the occurrence of target problem and the comorbid problem. In a

literature review, Gillberg and Billstedt (2000) reviewed the medical diagnoses which sometimes coexist with ASD, including Tuberous Sclerosis Complex, Fragile X Syndrome, and Partial tetrasomy. Comorbidities also included behavioural and motor control problems such as Obsessive Compulsive Disorder (OCD), ADHD, Catatonia Movement Disorder, Depression, eating disorders, and childhood schizophrenia (Gillberg & Billstedt, 2000). Kuusikko et al. (2008) also reported social anxiety in children and adolescents with AS, and the results showed that as age increased, behavioural avoidance and society anxiety increased as well. Other comorbidities of AS also include gender identity disorder, categorized by significant level of distress over discomfort at one's assigned sex, and having a strong desire to be cross-gendered (Gallucci, Hackerman, & Schmidt, 2005; Kraemer, Delsignore, Gundelfinger, Schnyder, & Hepp, 2005).

In terms of OCD, which is characterized by displaying repetitive behaviours in obsessions and compulsions (Robinson, 1998), Zandt, Prior, and Kyrios (2007) compared the levels of obsessive and compulsive behaviours between children with ASD (including children with AS), OCD, and a normal control group to determine the group similarities and differences. The issue of OCD as a secondary diagnostic of ASD was also addressed. The researchers found that in the ASD and OCD group, sameness behaviours and repetitive movements were at about the same levels. When each particular type of behaviour was examined separately, there were more compulsions and obsessions displayed in the OCD group. There was also an age effect in the OCD group, in which younger children tended to display more sameness behaviour, and older children usually showed more obsessions. This age effect did not apply to the ASD group. Zandt et al. (2007) concluded that it would be very hard to make a distinction between OCD and

ASD because they are very similar in characteristics of repetitive behaviour. Whether OCD should be a comorbidity of ASD is still questionable.

2.1.5 AS and Autism

Characteristics of AS and AD are often similar except for two major components – language delay and cognitive development. In order for an AS diagnostic to be made, an individual must not show significant delay in development of language, and a normal language development implies "non-echoed, communicative use of single words must be demonstrated by age 2 and meaningful phrase speech by age 3" (APA, 1994; Ozonoff et al., 2003, p. 6). If a child has displayed any language or cognitive delay, a diagnosis of autism will be made instead (APA, 1994). The DSM-IV (APA, 1994) also stated that in order to obtain a diagnosis of AS, the individuals must not meet the criteria for another PDD or Schizophrenia (Ozonoff et al., 2003).

The argument on whether AS should be treated as a separate diagnostic is controversial due to their similarities in symptoms. One year before Asperger illustrated his cases of 'autistic psychopathy', Kanner (1943) brought up the issues of infantile autism to which both descriptions were very similar in comparison (Wing, 1981). Due to this reason, until recently, some researchers fail to identify or to make a distinction between high-functioning autism (HFA) and AS. The two groups tend to be paired together as a single category when researchers report findings of research. However, it is very important to ensure such distinctions are made within the literature to avoid confusion if there are significant differences in symptoms or level of symptom severity between the two conditions. The normal IQ of individuals with HFA has also leads to the controversy on whether AS and HFA are the same. A major distinction between the two is the normal development of language or cognitive development with no significant delay in AS (APA, 1994). On the other hand, a study by Thede and Coolidge (2007) investigated

personality and neurobehavioural differences between children with HFA and AS and found more similarities than differences between the two groups' scores on personality scales. In comparison to a control group, both the AS and HFA groups showed significantly higher scores for both the Executive Function Deficit scale and the ADHD scale. The only difference shown between the two groups was the significant higher scores for the AS group on the Generalized Anxiety Disorder scale (Thede & Coolidge, 2007). More research is needed to determine if AS and HFA are the same.

2.1.6 Causes

The cause of AS is mainly biologically based and appears to be due to both genetics and/or neurology (i.e. neural connections, brain regions). Ghaziuddin (2005) performed a study on family histories of AS patients. Compared to a control group with autism, the first degree relatives of AS patients showed signs of AS, their relatives were also more likely to have higher rates of schizophrenia, depression and broader autistic phenotype. The author described broader autistic phenotype as "a lesser variant of autism" (Ghaziuddin, 2005, p. 3). Dorris, Espie, Knott, and Salt (2004) have demonstrated that siblings of children with AS showed poorer performance on a social cognition test (i.e. the Eyes Test) in comparison to a control group, suggesting that social cognition may have a genetic basis. In terms of brain structure, Stone, Baron-Cohen, and Knight (1998) have suggested that the frontal lobe of individuals with AS may be different from the 'typical' others. The authors found similar performance between participants with AS and patients with bilateral orbito-frontal (OFC) lesion on several ToM tasks. Individuals with AS had difficulties when asked to perform tasks that required more subtle social reasoning (Stone et al., 1998). Tani et al. (2006) examined neurological abnormalities in young adult participants with AS using a rating scale for clinical neurological abnormalities. When compared to a healthy

control group with no AS history, the AS group was found to differ in terms of gross and fine motor skills deficit, in which there were higher total scores for neurological abnormalities and scores of the neurological soft signs, that are anatomically and non-specific, which "characterized by abnormalities in motor, sensory and integrative functions" (Tani et al., 2006, p. 253). Neurological soft signs are believed to be responsible for dysfunction in sociopsychological because it was evident in the central nervous system (Quitkin, Rifkin, & Klein, 1976).

2.1.7 Effects/Impacts of As for Individuals with AS

Having a deficit in social communication is not easy for individuals in a society where interactions are important. Having AS would likely affect the person's quality of life and the way one perceives oneself. Hurlbutt and Chalmers (2004) examined adults with AS and their experience in the workplace using a qualitative method. Three themes were generated from interviews with six adults with AS: (a) adults with AS experience frequent unemployment and underemployment situations; (b) several factors affect employability, including social skills, communication, and sensory issues; (c) recommendations for aiding success in the workplace were also generated (Hurlbutt & Chalmers, 2004). Barnhill (2007) examined outcomes in adults with AS and identified characteristics that were associated with adulthood AS including employment issues, comorbid mental and physical health conditions, neurological and sensory issues (e.g., oversensitivity to touch and noise), possible problems with the legal system (e.g., crimes), mortality rates, and treatment problems such as their effectiveness (Barnhill, 2007). These examples demonstrate that individuals with AS often have a difficult time adapting to society, and their experiences do not seem to be positive.

2.1.8 AS Measures

Many of the instruments used to assess ASD such as the Childhood Autism Rating Scale (CARS; Schopler, Reichler, & Renner, 1988), the Autism Behaviour Checklist (ABC; Krug, Arick, & Almond, 1980), and the Gilliam Autism Rating Scale (Gilliam, 1995) usually include a category to distinguish AS (examples can be seen in Table A1 of Appendix A). However, screenings or diagnostic measures specifically for AS are still particularly new in the field. Although the 'Gold Standard' often used by practitioners and psychologists to make diagnoses are the DSM-IV-TR (APA, 2000) and/or the ICD-10 (WHO, 1993), the DSM-IV-TR (APA, 2000) is more commonly used in North America. Several other assessments exist, including the Autism-Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin & Clubley, 2001), the Adult Asperger Assessment (AAA; Baron-Cohen, Wheelwright, Robinson, & Woodbury-Smith, 2005), and the Asperger Syndrome Diagnostic Scale (Myles, Bock, & Simpson, 2001) (see Table A2 of Appendix A). The difference between screening tests and diagnostic tests should be identified: Screening tests are usually for casual, simple and rapid administrations while diagnostic tests are normally administered by trained professional and are used for building an in-depth profile of an individual (Stoez, Montgomery, & Hellsten, 2010).

2.1.9 Instrument Critique

It is not surprising that the measures of ASD that have been around for a long period of time generally seem to have more evidence of validity and reliability than newer instruments. However, the majority of the AS screening tests or instruments do not seem to take measurement theories into consideration before publication. Many of these popular measures lack psychometric justification outlining their practicality and appropriateness, as Stoesz et al., (2010) discussed in an evaluative paper of five AS instruments for adults. Stoesz et al. (2010) using the

Standards for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999) as a guideline found that the "accumulated evidence for normative information, reliability, and validity of each of these instruments are relatively poor" (p. 2). This section will examine the AS measures in detail, with specific attention paid to those assessments for adults. Measures discussed here include the (1) Ritvo Autism and Asperger's Diagnostic Scale (RAADS; Ritvo et al., 2008); (2) the Autism-Spectrum Quotient (AQ; Baron-Cohen et al., 2001); (3) the Adult Asperger Assessment (AAA; Baron-Cohen et al., 2005); and the (4) Australian Scale for Asperger Syndrome – Adult version (Meyer, 2000; Attwood, 1998). The Asperger Syndrome (and High Functioning Autism) Diagnostic Interview (ASDI; Gillberg, Gillberg, Wentz, & Rastam, 2001) and the Gilliam Asperger's Diagnostic Scale (GADS; Gilliam, 2001) will also be examined even though they were not developed specifically for adults.

2.1.9.1 RAADS. The Ritvo Autism and Asperger's Diagnostic Scale (RAADS) is a fairly new self-rating scale that was developed based on diagnostic criteria from the DSM-IV-TR (APA, 2000) and ICD-10 (WHO, 1993; Ritvo et al., 2008). There are 78 questions representing three domains: (a) social relatedness, (b) language and communication, and (c) sensorimotor and stereotypes. The response option of this scale is a four-point Likert-type scale. The items were written in first person perspective. Sixty of the questions were positively worded and the remaining 18 items were written negatively for the purpose of identifying 'normally expected answers' and to prevent response bias (Ritvo et al., 2008, p. 215). In terms of the development process of the RAADS, 100 items were initially developed; the items were then evaluated by three judges that were clinicians specializing in ASD. Two field trials were also conducted to

determine the quality of items. The scale was first administered to 32 participants (8 individuals with AS, 8 with AD and 16 normal individuals) for reviewing, the feedback reviewed was used for revision and deletion of items. Twenty new participants reviewed the items. Following this, 22 items were deleted due to redundancy, lack of clarity, or ambiguity after the two trials. The final items were assigned to domains and later reviewed by the same three judges. The RAADS takes less than one hour to complete.

The standardization process of the RAADS was conducted using a pilot study with 94 participants (17 with AD, 20 with AS, and 26 with other DSM-IV-TR (APA, 2000) diagnostic, and 41 normal individuals). The individuals with AD and AS were combined into one group, individuals with other DSM-IV-TR (APA, 2000) diagnostic and the normal people together formed a comparison group. The mean scores of the two groups were significantly different (F = 256.49, p = <0.0001) and revealed the cut off scores to determine whether a person has AS/AD. From examining the means, standard deviations, and ranges of scores from each of the diagnostic group demonstrated that a cut off score of 65 or below indicates a person is 'highly unlikely' to have AS or AD, and a score of 77 or above means it is 'highly likely' the person is having AS or AD (Ritvo et al., 2008, p. 217). For each of the questions, a post-hoc unequal variance t-test was done for pair-wise comparison between the two groups. It was found that the two groups differed significantly on 77 out of 78 questions (all $p < 0.002$). Internal consistency was computed for each of three subscales and Cronbach's alpha coefficients ranged from moderate to good (social relatedness, $\alpha = 0.86$; language and communication, $\alpha = 0.60$; and sensorimotor and stereotypies, $\alpha = 0.70$). According to Nunnally (1978, p. 245), a reliability coefficient of .70 is the minimum acceptable internal consistency value. Further analysis showed that removing two items from social relatedness would increase the alpha to 0.65, and removing one item from

language and communication lead to an alpha of 0.73. However, the authors decided to keep the items in the final version of the scale with no further explanation provided. Factor analyses were then conducted comparing AS + AD groups with the comparison group to determine the factor loadings of each subscale. Percentage of multidimensional variance for each set of questions ranged from 25.3% to 35.6%, and reliability ranged from 0.761 to 0.909. An analysis of variance (ANOVA) showed a p-value of less than 0.0001 for each of the subscales.

Ritvo et al. (2008) noted that a larger sample size is needed for some of these results to be more precise. For example, a more exact cut off score may be generated with greater sample size. At this current level, a score of 64 or less indicates an individual is unlikely to have AS or AD, a score of 77 and above means the person is very likely to have AS or AD. However, because no one in the sample scored between 65-77, a greater sample size is needed to bring this gap closer.

Although the authors did attempt to collect content validity evidence by involving expert judges, the target population (AS and AD people) and healthy individuals, one problem was the small number of judges (3) they had. If one or more expert judges had been added, more trustworthy information may have been collected. Haynes, Richard, and Kubany (1995) suggest that having more than five judges would increase the detection of rater outliers and allows for the removal of any aberrant judges if necessary. Furthermore, if judges' agreements and other analyses of content validity were to be conducted, more judges are needed. The authors did not mention how the rating results were analyzed.

In terms of internal consistency, an explanation should be provided as to why the authors decided to keep the three items which if omitted could bring the internal consistency higher. Test-retest reliability could also be assessed in order to gain more evidence of reliability.

Although the scale was built based on criteria on the DSM-IV-TR (APA, 2000) and the ICD-10 (WHO, 1993), convergent and divergent validity evidence could be collected by using scales that are meant to measure similar or different constructs (DeVellis, 2003). After examining the items on the scale, a few questions were double-barreled with '*and*' conjoining two possible responses. Some items also appear too long to easily read (e.g., "If I am in a place with many smells, textures to feel, noises, or bright lights; I can get overwhelmed with sensations and feel panicky, anxious, or frightened."). Both double-barreled items and items that are too long may affect the reliability of the measure itself, thus causing complexity and lesser clarity of the question (DeVellis, 2003), because people may misread the questions and agreed to one part of the question but not another. It would be useful to include people familiar with instrument development as judges to review the items to ensure they are reasonable. Ritvo et al. (2008) noted that the RAADS does not distinguish individuals with AS from individuals with AD.

The RAADS appears to be lacking a few psychometric properties (i.e. test-retest reliability, evidence of concurrent validity and divergent validity). Content validity was conducted with only three judges and thus this evidence can be considered weak. At this time, RAADS should not be used as the sole tool to assist practitioners in making a diagnosis or even making classifications for individuals suspected of having AS or AD. A larger sample size should be used if factor analysis is conducted. Items should be revised again, and the instrument would be more useful if it could distinguish between AS and AD. If a diagnosis is to be made, the RAADS should not be used alone and other diagnostic measures are needed.

2.1.9.2 AQ. The AQ was developed because at the time of development there was no short and self-administered instrument that measured "the degree to which an adult with Asperger with normal intelligence has the traits associated with the autistic spectrum" (Baron-

Cohen et al., 2001, p. 6). The AQ contains 50 questions in total measuring five domains (social skill, attention switching, attention to detail, communication, and imagination) with 10 questions in each domain. All abnormality items would have a score of 1, and response bias is prevented with half of the items worded to generate 'agree' response and the other half with a 'disagree' response. The authors mentioned that the items were based from a 'triad' of ASD symptoms and also from cognitive abnormality that has been found to be related to autism. Multiple pilot studies were conducted over the years to modify the instrument. An earlier version of the AQ was interview based, and it was tested with the AS/HFA adult population with age-matched controls. Although there was some concern whether people with AS/HFA would have problems comprehending the instrument, this seemed to be unfounded because the target population of AQ are people who are higher functioning, therefore individuals should be able to read and respond to the instrument (Baron-Cohen et al., 2001). A later pilot study was also conducted to examine the concern that people with AS/HFA might not be able to judge their own behaviours because of their problems in subtle mind reading. Results have shown that these individuals were able to report their preferences appropriately (Baron-Cohen et al., 2001). However, questions were asked in terms of their preferences (to guard against false negative stated by the authors), instead of keeping the original item forms of the AQ. This pilot study was not described well enough, and there was no comparison group (or control) group to correlate their results with, for example, parents' rating results could also be used as comparison. Therefore, conclusions drawn from these pilot studies should be done with caution.

Baron-Cohen et al. (2001) conducted a study of AQ with 58 adults with HS/HFA of normal IQ range, 174 adults from random selection, 830 Cambridge University students, and 16 winners from a UK Mathematics Olympiad. Results have shown that the AS/HFA group

generally scored higher than the control, randomly selected adult group. Within the control group, males tended to have a higher score than females, but this difference was not found in the AS/HFA group. In examining whether university students would score differently than the control group due to their possible differences in education level and intelligence, it was determined that they were very similar in scores. Dividing the student group into science (e.g., physical sciences, mathematics, and biological sciences), humanities (e.g., languages, law, and history), and social sciences field (e.g., anthropology, management and geography), students in science scored higher than humanities and social sciences. Baron-Cohen et al. (2001) suggested that this difference provides further indications about the correlation between autism and skills in sciences and mathematics. Breaking the field of science even further, the authors found that students in mathematics "scored higher than engineers, physical and computer sciences, who scored higher than medicine and biology" (Baron-Cohen et al., 2001, p. 10). To examine this idea even further, the winners of UK Mathematics Olympiad were compared with the male humanities students, and as predicted, the winners in mathematics had higher AQ scores (Baron-Cohen et al., 2001).

Psychometric properties of the AQ were also determined in the same study by Baron-Cohen et al. (2001). Test-retest reliability was conducted by selecting only 17 participants from the student group to complete the AQ two weeks later. Scores were not significantly different ($t(16) = .03$, $p = .75$), correlation of scores between the two periods were 'strongly correlated' as explained by the authors ($r = .7$, $p = .002$). A correlation of .7 is normally considered as acceptable (Nunnally, 1978), but not to the level of 'strongly correlated'. This sample is also of concern, as the AQ is targeted for HS/HFA individuals, having the HS/HFA group in the stability study would seem to a more convincing test of how reliable the test is for that specific

population. The sample size (n = 17) used to calculate the test-retest reliability is too small and a larger sample would allow for more convincing evidence. Comparing the parent report and the self-report of AQ revealed that parent scores were higher than the self-report scores, in which scores of the AS/HFA self report would be considered more conservative as noted by the authors.

Examinations of items by Baron-Cohen et al. (2001) found that two items tended to have higher scores in the control than the AS/HFA group. Internal consistency was calculated to determine how similar the items were in each domain, Cronbach's alpha coefficient ranged from 'moderate to high'. However, the majority of subscales (four out of five subscales) had reliability coefficients below the acceptable Cronbach's alpha of .7. As the coefficients ranged from .63 to .77, the authors should consider revising some of the items. The study also attempted to determine a cut off score for the AQ by applying a rule that "a useful cutoff would discriminate the groups with as many true positives and as few false positives as possible" Baron-Cohen et al. (2001, p. 12). How exactly this process was done was not explained, but a score of 32+ was set due to 79.3% of AS/HFA group reaching those scores versus 2% from the control group. According to the authors, the same cut off score can be applied to females because 92.3% of AS/HFA females scored 32+, compared to 1% of the control females. The male results were not discussed. Among the participants who scored an AQ score of 32+ in the control group, the researchers had these individuals (n = 11) assessed through a clinical interview. Seven participants actually met the criteria of AS/HFA but a diagnosis was never received. From the results of this study, Baron-Cohen et al. (2001) concluded that the AQ has 'reasonable construct validity' based on what they found (p. 14). The authors did not consider any additional comparisons such as computations using analysis of variances (ANOVA) comparing control

versus AS/HFA group, students from different departments (scientists versus nonscientists), and students within different types of sciences as evidence of construct validity. However due to the correlation between strong mathematics skills and ASD as discussed by the authors (Baron-Cohen et al., 2001), and in theory, those with AS/HFA should score higher than the control group, this information could be considered some evidence of construct validity. Face validity was also claimed to be reasonable due to the result that AQ scores of the AS/HFA group were significantly higher than the control group, but this result is more appropriately classified as construct validity. Baron-Cohen et al. (2001) also noted that the purpose of the AQ is to distinguish those who have AS/HFA, but it is not meant to make a diagnosis because it does not take into account whether these characteristics are affecting an individual's life to any great extent.

The AQ was also examined in another study by Woodbury-Smith, Robinson, Wheelwright, & Baron-Cohen (2005). One hundred adults being referred to an Asperger Syndrome service participated in the study to test the AQ's ability to distinguish between AS and HFA individuals. The DSM-IV (APA, 1994) was used as a comparison of the AQ's accuracy. Diagnosis on the DSM-IV (APA, 1994) was made by clinicians conducting an interview with the participants and their informants upon completion of the AQ. Discriminant validity was shown when the AQ was able to distinguish individuals with a clinical AS/HFA diagnostic and those who showed similar characteristics but not severe enough to be diagnosed. The authors came to this conclusion by examining measures using parametric statistics, and by looking at the area under a few receiver operating curves (ROC). The area under a ROC usually represents the accuracy of a test. It indicates the "probability that a randomly selected 'true-positive' individual will score higher on the test than a randomly selected 'true-negative' individual" (Woodbury-

Smith et al., 2005). In this study, an area of .78 (std. err. 0.06, 95% CI 0.7-0.9) was found, in which it was considered as moderate by the researchers because this number represents the accuracy of the AQ. The ROC was also used in the study to set a more precise cut off score than the suggested 32+ from before. A score of 26 was decided on as a better cut off score. At this level, about 83% of the people received the correct diagnose (sensitivity of 0.95, with specificity 0.52, positive predictive value 0.84, and a negative predictive value of 0.78; Woodbury-Smith et al., 2005). One limitation noted by the authors was the notion that clinicians doing the interview were not blind to the people's AQ scores. Another possible limitation suggested was the influence of confounding factors that could affect the AQ scores. This study did not take into account other mental health problems, and it is a huge factor that could possibly affect AQ scores when both the mental health problems and AS do display similar symptoms (e.g., schizophrenia).

More validity and reliability evidence is needed for the AQ. For example, measuring concurrent and divergent validity would definitely provide further information on how accurate the AQ is. Conducting test-retest reliability with a larger AS population would be beneficial. In terms of the development process of AQ, Baron-Cohen et al. (2001) should provide more detail on what is considered as *trait* symptoms that the items were based on.

2.1.9.3 AAA. The AAA is a computerized assessment using Microsoft Excel designed for practitioners to identify adults with AS from adults with autism (Baron-Cohen et al., 2005). Although Baron-Cohen et al. (2005) claimed there was no measure at that time to assess AS and HFA in adults, and AAA would serve that purpose, Wing and Gould (2006) pointed out that there were existing scales like the ASDI and the Diagnostic Interview for Social and Communication Disorders (DISCO) with the same abilities. Diagnostic criteria were claimed by the authors to be stricter than the DSM-IV (APA, 1994; Baron-Cohen et al., 2005). The AAA

does attempt to distinguish between AS and HFA by simply looking at the language delay component of a client. Individuals with language delay are categorized as HFA and individuals without language delay are classified as belonging to the AS group (Baron-Cohen et al., 2005). The AAA consists of five sections (A to E) with items built from the DSM-IV (APA, 1994) and additional factors that the authors believe to be important but are not included in the DSM-IV (APA, 1994). The five sections are (a) Qualitative impairment; (b) Restricted repetitive and stereotyped patterns of behaviour, interest and activities; (c) Qualitative impairments in verbal or non-verbal communication; (d) Impairments in imagination; and (e) Key pre-requisites. Section A and B have four items from the DSM-IV (APA, 1994) and one additional item; none of the items from Section C are from the DSM-IV (APA, 1994) criteria for AS but from part of the DSM-IV (APA, 1994) criteria for AD. They are worded so that they are "appropriate for diagnosing adults with AS" (Baron-Cohen et al., 2005). Section D has a similar nature to Section C, but with one item from the DSM-IV (APA, 1994) criteria of autistic disorder, and two additional items. In order for an AS diagnosis to be made, the individual must meet all the pre-requisites in Section E, at least 3 symptoms in Section A-C each, and one symptom from Section D. The AAA requires a minimum score of 10 out of 18 to classify an individual as having AS.

In terms of the development of the AAA, there was no mention as to how additional items were created and the selection of pre-requisites. There was no standardization performed on the AAA, and there was no report of reliability evidence of the AAA as well (Stoesz et al., 2010). A validation study by Baron-Cohen et al. (2005) demonstrated that three out of eight individuals classified as normal by the AAA actually had a diagnostic of AS by the DSM-IV (APA, 1994). One cannot be sure whether the AAA has a tendency to misclassify, or if it was being conservative as mentioned by Stoesz et al. (2010). The AQ and the Empathy Quotient

(EQ) are two measures that link the AAA together, and clients are asked to complete both questionnaires before the scoring of the AAA is conducted. No discussion was provided as to why the two measures (i.e., AQ and EQ) were chosen, but it should be noted that the same research group developed them both. Thus the AQ and EQ are used as screening questionnaires prior to the use of the AAA. The characteristics of AQ were discussed above. The EQ is an instrument containing 60 questions (40 items to measure empathy and 20 filler items) that assess empathy in individuals. People with AS are often found to have lower levels of empathy than the normal population (Baron-Cohen et al., 2005). Responses on the AQ and the EQ are entered into the computer template and the resulting scores are calculated and entered on the first page of AAA. Appropriate responses on the AQ and EQ that are relevant are also entered into their corresponding domains of the AAA. Since the AAA is dependent on the AQ and the EQ, the psychometric properties of these measures must also be reasonable in order for valid and reliable interpretations of the AAA scores to take place. However, the authors did not discuss the EQ further. Therefore, the EQ will be critiqued next.

According to Lawrence, Shaw, Baker, Baron-Cohen & David (2004), interpretations of the scores resulting from the EQ were validated using 90 individuals with AS/HFA and a control group of 197 normal participants. Results have shown that the EQ is able to reliably distinguish between the AS/HFA and the control group and that over a 12 month period, the EQ demonstrated high test-retest reliability (Lawrence et al., 2004). However, the authors did not provide any statistical numbers or results to back-up these claims. Content validity was conducted by six judges rating the items of the EQ to the definition of empathy. Baron-Cohen and Wheelwright (2004) have demonstrated the results that the judges were able to identify correctly the filler items to which they were unrelated to empathy, and all other empathy related

items were rated as relevant by the judges. It was also reported that, "the probability of obtaining such agreement on each item by chance is $p < .003$" (Baron-Cohen & Wheelwright, 2004, p. 168). All relevant items were agreed to by the judges as being related to the authors' definition of empathy and filler items were all categorized as not belonging to the construct of empathy. However, the scale of the judges' ratings was based only on yes and no responses. A Likert type scale might provide a better idea of the relevancy of each item. Convergent and divergent validity evidence for the EQ were accumulated in several studies. Results of the AQ, the Friendship Quotient (FQ), and the EQ were compared. As predicted by Lawrence et al. (2004), scores on the AQ scale were inversely correlated to scores on the EQ scale ($r = -0.56$, $p < .0001$) and positively related to scores on the FQ scale ($r = 0.59$, $p = < .001$). The FQ scale is "a 25-point self-report scale assessing reciprocity and intimacy in relationships" (Baron-Cohen & Wheelwright, 2004, p. 166). Friendship is a quality that is also expected to be relatively low in people with AS. Additional validity studies were also carried out by Lawrence et al. (2004) by comparing the EQ scale with the Eyes test (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), and the Interpersonal Reactivity Index (IRI; Davis, 1980). These instruments were developed as measure of people's capability to read emotions through the eyes, and to assess empathy respectively. Lawrence et al. (2004) reported a 'modest positive relationship' between total EQ score and the Eyes test score ($n = 48$, $r = 0.294$, $p = .033$; p. 913). Results have also shown that scores on the IRI were correlated moderately with scores on the EQ scale for two of the four subscales ($p < .05$), and the relationship for the remaining two subscales were found to be insignificant ($p > .05$; Lawrence et al., 2004). Test-retest reliability was also carried out by Lawrence et al. (2004) with 25 participants after 10-12 months of initial testing. Test-retest reliability of the EQ scale was good, with a correlation coefficient of $r = .835$ ($p = .0001$).

However, since empathy level of an individual should remain stable over time, the time frame between the test-retest could be stretched to about two years to ensure the high reliability was not due to participants remembering the questions. In addition, when the time has been extended, the number of participants should also be increase due to possible participant dropout. Along with Lawrence et al. (2004), Baron-Cohen and Wheelwright (2004) also reported a fairly high coefficient of reliability, $r = .97$ ($p < .001$) for the EQ. Internal consistency was conducted by Lawrence et al. (2004), in which the correlations between items were determined. Items that had low correlations with other items (< 0.2) were recommended to be discarded. Another factor that was also examined was the participants' scores on the Social Desirability Scale (SDS), or their tendency to respond to questions in socially desirable ways (Lawrence et al., 2004). Correlations between the SDS and the items were calculated. Items that were found to generate a high social desirable tendency were marked for deletion. Using this process, only 28 items across the three factors were recommended to be kept.

Overall, the EQ scale should be revised with poorly performing items dropped. Concurrent and divergent validity should also be conducted with measures that are not developed by the researchers involved with the EQ to generate more objective independent evidence.

2.1.9.4 The Australian Scale for Asperger Syndrome – Adult Version. This measure was 'created' by Meyer (2000), where it was modified from the 'upgraded' Attwood (1998) version of the original Australian Scale for Asperger Syndrome by Garnett and Attwood (1995). This version is unapproved and is available on the Internet only as a draft. The instrument consists of 80 questions across five subscales and there is no estimated time on how long it should take an individual to complete. Items are worded in a third person perspective and the questionnaire is not designed for self-reporting. Some of the questions were developed by Meyer in addition to

the original scale to address the challenges adults with AS would often encounter (Meyer, 2000). How exactly the items were written was not discussed. The five subscales are (a) Social and Emotional Abilities; (b) Communication Skills; (c) Cognitive and Executive Function Skills; (d) Somatic, Motoric, and Presentation Issues; and (e) Other Characteristics. Items were developed based on the original version for children, and items were worded to match adult population (i.e. change of wording). A 6-point Likert scale was used as a response format similar to the original form. The scoring method was not described, but Meyer (2000) directed the users to page 20 of 'Attwood's book', but reference to the book was not provided anywhere on the website. Meyer (2000) also noted that if a person is experiencing the majority of the symptoms checked as 'yes', it does not automatically mean a person has AS because it could possibly be HFA instead.

The original version of the Australian Scale for Asperger Syndrome can easily be found online. However, there seems to be no published literature that examines the validity and reliability of the original scale. Howlin (2000) did a brief review of this scale, and concluded that no psychometric properties have been reported. After reviewing the items in the adult version, the questions seemed to be long in length. Many questions contained double-barreled phrases were also found and a few questions could be found being grouped within one question. For example, question 20 asks:

> Does the person have an unusual tone of voice (monotone), sing-song an "affected" foreign accent, unusual inflections, prosody, and other oddities of fluid speech? Does he speak with an unusually loud or soft volume level? Does he use changes of tone, inflection, or volume levels appropriate for different levels of conversational formality, location, and topic choice?

Several questions were clustered into one item. To improve measurement, the items should be made separate.

Some questions were unclear and lacked explanations. For example, in the first question "Does the person lack understanding of how to play adult games with others? For example, unaware of the unwritten social rules of leisure and recreation." What exactly is classified as 'adult games' was not explained. The writer did not appear to do a good job of modifying the original questions written specifically for children into an adult context. It is very important for evidence of the validity and reliability of this scale to be accumulated before interpretations and decisions can be made. At this stage, it seems inappropriate to be widely publishing this scale on the Internet for the public because it has the potential to provide misleading results.

2.1.9.5 ASDI. The ASDI (Gillberg, Gillberg, Rastam, & Wentz, 2001) is a 20-item interview consisting of six constructs (or criteria as described by the authors) for the practitioners to use. The authors saw the need for constructing the interview because they noticed the lack of diagnostic interviews made specifically for individuals with AS/HFA, and screening tools are not able to serve the purpose of providing a detailed clinical evaluation (Gillberg et al., 2001). The six constructs listed by the authors were: (a) Severe impairments in reciprocal social interaction (extreme egocentricity; four items); (b) All absorbing narrow interest pattern(s; three items); (c) Imposition of routines, rituals and interests (two items); (d) Speech and language peculiarities (five items); (e) Non-verbal communication problems (five items); and (f) Motor clumsiness (one item). Originally there were three possible ratings on each item (1 = does not apply, 2 = applies sometimes or somewhat, 3 = definitely applies), but for unexplained reasons, the newer version now utilizes two possible ratings (0 = does not apply, 1 = applies to some degree or very much) on each of the 20 items. This interview is to be used by clinicians familiar with AS and

related ASD, but it is not necessary for clinicians to reach a specific level of expertise (Gillberg et al., 2001). In addition, the researchers provided no specific age range for the target population, but it was mentioned that ASDI is for adolescents and young adults.

The interview was not developed based on the DSM-IV-TR (APA, 2000) nor the ICD-10 (WHO, 1993) due to the reason that Hans Asperger's cases of AS did not meet the diagnostic criteria of the DSM-IV (APA, 1994) and the ICD-10 (WHO, 1993) for AS. The DSM-IV-TR (APA, 2000) also requires normal development of language in order to meet the requirement of AS, however, Gillberg et al. (2001) noted that Asperger's cases did not mention language delay. Therefore, the items were developed based on the authors' experience in the field and also on Hans Asperger's reported cases. No criteria was provided on how items were developed in which there were no definitions provided for AS and HFA.

Gillberg et al. (2001) conducted a preliminary study on the ASDI with 24 participants with different kinds of neuropsychiatric disorders (seven are normal control) and their first-degree relatives. A diagnosis was conducted for the individuals before the study by two neuropsychiatrists or one neuropsychiatrist plus a neuropsychologist familiar with ASD. In the inter-rater reliability study, interviews with the relatives of 20 individuals with AS (n = 8), atypical autism (n =2), OCD (n =2), multiple personality disorders (n =1), and control individuals (n =7) were conducted by two neuropsychiatrists. Participants were asked to complete a neuropsychological test and diagnoses of AS were made based on the DSM-IV (APA, 1994), the Gillberg and Gillberg (1989), and the Szatmari (Szatmari et al., 1989) criteria. The neuropsychiatrists took turns interviewing the relatives while completing the ASDI, and the other acted as observer, independently filling out the ASDI. The relatives and the interviewer were both blind to previous diagnostic results. The percentage agreement was calculated as 96%

with a kappa of .91, which is considered a good level by Gillberg et al. (2001). Intra-rater reliability was also examined with four additional participants' (n = 24) relatives of AS (n = 2), atypical autism (n = 1) and ADHD (n = 1) 10-15 months after the initial interview. This retest of ASDI was completed by one of the two neuropsychiatrists interviewing the relatives again. Why the four additional cases were added was not explained, and these individuals do not seem to have an initial ASDI score for comparison. It was not sure how the calculation was conducted for the four extra cases if the initial score was not available. Percentage agreement was calculated as representation of test-retest reliability, 97% agreement was found for this study with a kappa of .92.

In terms of validity, there was no content validity evidence shown in the study. Items from the instrument may be biased towards the researchers' subjective view on how they believe the construct should be. Gillberg et al. (2001) attempted to show concurrent validity by demonstrating individuals with a clinical diagnosis of AS also met most of the criteria in ASDI. However, the authors did not compute any form of correlation between the scores of the measures used in diagnosis and the ASDI. Only a brief description was discussed comparing the numbers of criteria scored on the ASDI by an individual who received a diagnostic of AS, with individuals who did not receive a clinical diagnostic of AS.

After examining the item quality of the ASDI, it can be seen that there are items with double-barreled phrases. Some of the items were written in a way that a 'yes' response would imply either one case or the other. These sentences often contained the conjunction 'or' to include two possible responses. Some terms were not clarified, and thus participants might not answer these questions correctly. For example, the item: "Was his/her language development delayed?", what was determined as delayed was not described, and therefore answers were solely

based on interviewees' own interpretation. Another example was the item: "Is his/her gaze stiff, strange, peculiar, abnormal or odd?" How many first-degree relatives, especially parents would label their child's gaze 'abnormal'? Another interesting point to note is the representation of the population in the study. The individuals with neuropsychiatric disorders aged from 6 to 55 years, which did not match the intended population of the ASDI (adolescents and young adults). In addition to the small sample size, the results of this study cannot be generalized to the ASDI target population. Overall, more research is needed to conclude whether the ASDI is a useful instrument due its lack of advanced analysis. In addition, the ASDI was not made to differentiate between AS and HFA (Stoesz et al., 2010), and therefore it would be harder to pinpoint whether an individual is potentially having AS or HFA. More evidence of reliability (e.g., internal consistency) and validity (e.g., content validity, concurrent validity, and divergent validity) is needed for the ASDI as well. In this study, the authors claimed that the ASDI has good validity, but we can see that there was little to poor validity evidence shown. Although much attention was paid to the reliability of the measure, reliability evidence is not the only important psychometric quality

2.1.9.6 GADS. This instrument requires purchase and it is not available on the Internet. Therefore, the following review and information of the GADS is based on reviews and critiques done by Campbell (2005), and Stoesz et al. (2010). The GADS is a norm-referenced rating scale that contains 32 items from four domains assessing individuals between 3 to 22 years. The four subscales are: (a) Social interaction; (b) Restricted patterns of behaviour; (c) Cognitive patterns; and (d) Pragmatic skills. Items were built based on a literature review of AS, the DSM-IV-TR (APA, 2000) criteria, the ICD-10 (WHO, 1993) criteria, and other instruments that assess AS. There are five purposes of GADS stated in the manual: (a) To identify individuals with AS; (b)

To assess individuals who show unique behavioural features; (c) To document behavioural progress; (d) To target goals for IEP (Individualized Education Program); and (e) For research use (Campbell, 2005, p. 27). Respondents of the GADS can be anyone who has had continuous contact with the suspected person for at least two weeks. There is also an additional Parent Interview Form to assess the history of cognitive and language delay of the individual. The GADS is scored by summing the raw scores to create four subscale scores. Percentiles are also calculated. Adding the four subscale scores will create a total score called the Asperger's Disorder Quotient (ADQ) in which it is a standard score with a mean of 100 and a standard deviation of 15. The GADS only takes about 5-10 minutes to complete. An ADQ score of 80 or higher indicates a very likely diagnosis of AS.

Originally there were 70 items developed which were later reduced to 32 items allocated across four subscales using data analysis. However, Campbell (2005) pointed out that there was no discussion on how the final 32 items were selected and there was no discussion on which data analyses methods were used to group these items into subscales. Norming and standardization of the GADS was conducted with 371 people (age from 3-22 years, 85% male) diagnosed with AS. However, Stoesz et al. (2010) pointed out that only the information on the United States participants were provided in the manual even though the normative data was done in the United States and seven other countries. According to Campbell (2005), school professionals and parents (contacted through the Internet) of individuals with AS were asked to complete the GADS. It was found that there were no sex and age differences within the sample, thus cumulative frequency tables were used to create the subscales and ADQ scores. A Cronbach's alpha of .87 was calculated based on 360 individuals from the standardization sample. The coefficients ranged from .70 to .81 for the subscales. Using 10 teachers' ratings of their AS

students, a temporal stability reliability was found to be .93 over a two-week period. However, it should be noted that a two-week period for temporal stability reliability is too short as teachers may still recall their ratings of the students from two weeks ago. The time period should be expanded to at least one month or more. In addition, more teachers should be included to ensure the accuracy of the reliability scores. On each subscale, the test-retest reliability was moderate, ranging from .71 to .77. Inter-rater reliability was .89 between teachers and parents of 16 children (10 with AS). Reliability of the GADS scores seems to be good.

Campbell (2005) believed that the GADS had content validity based on the test developer's discussion of the DSM-IV-TR (APA, 2000) and the ICD-10 (WHO, 1993) for item development. The GADS was able to distinguish the AS individuals with others of different diagnoses (i.e. ADHD, learning disabilities, and mental retardation). However in my opinion, one cannot be sure of the quality of the items themselves (e.g., reading level, presences of double-barrels, understandability, and wording) without obtaining judgments from experts – a major component of content validity. Evidence of construct validity was also claimed due to the strong internal consistency values found and that there were differences in scores between AS and other disabilities.

In another evaluative study, Stoesz et al. (2010) discussed further validity evidence of the GADS. With a sample of 50 children, there was only a "positive moderate" relationship (p. 13) found between the GADS and the Gilliam Autism Rating Scale (GARS; Gilliam, 1995). The correlation was .58, $p < .01$. If the two instruments were measuring the same construct, their correlation should be higher than a moderate level. Overall, the GADS appears to be a reasonable instrument in identifying individuals with AS. More evidence of validity should be collected especially with respect to item development and item selection process. It would be

preferable if more validity properties such as criterion-related validity were also examined. Like Campbell (2005) pointed out, the sample size in the standardization study appropriately represented the target population. However, the test author did not confirm the diagnoses of their sample, which may affect the representation of the AS population from the sample, and we do not know whether the GADS is accurate in identifying true individuals with AS. Furthermore, there were only a few adults employed in the standardization study and this leads to a generalization problem for the use of GADS in the adult population (Stoesz et al., 2010).

2.1.9.7 Overall Quality of the Scales. As can be seen from the critique, many of the available scales lack basic psychometric properties such as validity and reliability. Some items were not well written. Many of the developers even failed to provide enough information on the measures for others to evaluate it. The only instrument that seemed to possess psychometric evidence is the GADS, but its target population ranged from only age 3 to 22. If a diagnosis is to be made for adults, it is strongly recommended that the DSM-IV-TR (APA, 2000) or the ICD-10 (WHO, 1993) should be used. If a self-screen is to be used, then there is currently no existing quality measure. If an adult is looking for a self-screen, the only available adult self-screen for AS/HFA is the RAADS or the AQ, but the psychometric qualities of both are still lacking and the RAADS may require further modifications. At this point, a better adult self-screen instrument would be preferable. The current study aims to develop items and evaluate the content validity of these items as an important first step in developing a new AS screening instrument.

2.2 Scale Development

Developing a useful instrument requires more work than just writing the items and distributing the test to the target populations. Many researchers often overlook factors such as writing good items, standardizing the test, and accumulating evidence of reliability and validity

for the interpretation of the test scores. Measures developed with little or no attention paid to these factors may result in a measure that is not useful in answering the researchers' questions. Hinkin (1995) pointed out that scholars are often embarrassed realizing that results gathered from their measures are inconclusive, and little is actually learned about the research question. If the poorly developed test is a high stakes test, it can affect test takers' designation, and decisions made based on this test will therefore be unfair and may lead to negative consequences. At this time, the main standard for test development is the *Standards for Educational and Psychological Testing* (AERA et al., 1999). This standard would also provide a more unified benchmark for test development and evaluations. Some researchers also suggest guidelines or rules for test developers to follow. Hinkin (1995) suggested three stages for test development: (1) item generation; (2) scale development; and (3) scale evaluation. DeVellis (2003) suggested a more detailed guideline for instrument development including eight steps: (1) determining clearly what it is you want to measure; (2) generating an item pool; (3) determining the format for measurement; (4) reviewing the initial item pool by experts; (5) considering inclusion of validation items; (6) administering items to a development sample; (7) evaluate the items; and (8) optimize scale length. The focus of this research study is step 2 through step 4. Items were written to develop a new AS self-screening instrument for adults, the items were sent out to expert judges for evaluation – a method of determining evidence of the content validity of a measure. This study also explored the question "Which method(s) are most efficient and useful in assessing expert panel member's judgments?"

2.2.1 Item Writing

Well-written items are very important to the applicability of an instrument. Frey, Petersen, Edwards, Pedrotti, and Peyton (2005) analyzed 20 classroom assessment textbooks and summed

up the recommendations within each with a guideline of 40 ‘rules’ most commonly seen in the textbooks. Four item writing blunders that have a direct connection with the validity of an instrument were the focus of these rules: (a) potentially confusing wording or ambiguous requirements (double-barrels), (b) guessing of answers, (c) test-taking efficiency, and (d) the control for testwiseness (Frey et al., 2005). Although some of the rules are irrelevant to this study as the 40 rules were developed for educational purposes, the 40 rules will be used as a reference when developing the items for this study (see Appendix B). DeVellis (2003) and Rogers (2010) also provided some suggestions for item writing. For example, ‘exceptionally lengthy items’ should be avoided; the level of reading difficulty should be considered; avoid the inclusion of double-barreled items and ambiguous pronoun references; and the use of positively or negatively worded items should take into considerations (DeVellis, 2003).

2.2.2 Reliability and Validity

Reliability and validity are two major concepts test developers need to pay attention to when developing an instrument, and they determine the usefulness and applicability of a measure. Although validity, specifically content validity is the center of this research, reliability is also described because reliability and validity are often paired together in research due to their importance. It should also be mentioned that reliability does not equate to validity. An instrument demonstrating evidence of one of these psychometric characteristics does not imply the presence of the other. Reliability can be present without validity but to be valid, there must be evidence of reliability.

2.2.2.1 *Reliability*. One question we would ask in terms of reliability, is that if we were to send out the same test to the same group of people, how similar (or consistent) would the test results be? Accuracy level of a test increases when a test can generate consistent scores.

Reliability of a scale is "the proportion of variance attributable to the true score of the latent variable" (DeVellis, 2003, p. 27). When a test score is obtained, it is assumed that this test score (also called an observed score) has both a true score and an error score associated with it (Rogers, 2010; Traub & Rowley, 1991). According to Cronbach (2004), "all psychometric theory of reliability pivots on the concept of true score" (p. 400). A true score is the test taker's true ability, it is the average of an individual's score over repeated testing infinite number of times (Cronbach, 2004). On the other hand, the observed score is often affected by the error score because true scores are assumed to be the same for the test taker over all occasions (Traub & Rowley, 1991). An error score represents the part of the observed score that is "unsystematic, random, and due to chance. It is the accumulating effects of all uncontrolled and unspecified influencing factors included in the test score" (Harvill, 1991, p. 181). Errors are represented by the variance or standard deviation (square root of variance) of the scores in a distribution when repeated measures are done, this standard deviation is also called the standard error of measurement (Traub & Rowley, 1991). A smaller error score is desirable as it indicates a closeness of true score and observed score, meaning the reliability level of the test is greater. Several ways of determining reliability are examining test-retest reliability, alternative forms of reliability, split-half reliability, internal consistency, and inter-rater reliability.

Of all the methods, often found in the literatures are test-rest reliability and the internal consistency. Test-retest reliability is determined by calculating the correlation when the same test is administered to the same group of people at two different times. Usually the correlation is higher when time 1 and time 2 of test taken is within a shorter period of time (Trochim, 2006), but too short a gap and people might still remember their answers from time 1. However, if the time gap is too long, then people's ability or knowledge may differ (e.g., a student may learn

more about a particular topic as the school term progress). If the purpose of a test is to determine student's knowledge on a particular topic, then its test-retest reliability would be affected when the time period between the two administrations was long. On the other hand, if an instrument is used to measure a certain disorder, and it was expected to be persistent within an individual over time, a high reliability score would indicate that the items measuring that disorder are reliable even when the time between administrations were long. Internal consistency concerns the homogeneity level of test items (DeVellis, 2003). It is determined by calculating the level of correlation between items in a measure. There are a few ways of determining internal consistency. Three popular ways for calculating the coefficient alpha are: Cronbach's alpha, the Kuder-Richardson formula 20 (KR-20), and the Kuder-Richardson formula 21 (KR-21). KR-20 and KR-21 are mainly used for dichotomously scored items and Cronbach's alpha is mainly used for polytomously scored items. The follow is the formula for Cronbach's alpha:

$$\alpha = k / (k-1) [1- (\Sigma S^2_i / S^2_t)] \quad (2.1)$$

where k is the number of items, ΣS^2_i is the sum of variance for separate test items, and the variance for the set of student total scores is represented by S^2_t (Frisbie, 1988).

2.2.2.2 Validity. Validity provides us with the idea of how much a test is actually measuring what the test developer intends it to measure (the construct). For example, an anxiety test that aims to determine test takers' level of anxiety in a particular situation should only contain items that are measuring anxiety, but not how much the test takers can understand the words on each question. Messick (1990) provided a definition of validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessments" (p. 1). Another way to describe validity is that it "concerns whether the variable

is the underlying cause of item covariation" (DeVellis, 2003, p. 49). This means if items in a test are measuring what they should be, they will change (vary) according to how the construct is defined. Validity can be categorized into four types including content validity, criterion-related validity, construct validity, and consequential validity. The different types of validity are not mutually exclusive and can be used according to the need of the evaluator. An argument made by Messick (1991) was that all these 'types' of validity should be considered as a form of construct validity and all are important in contributing to the validity of a test.

2.2.2.3 Content validity. According to DeVellis (2003), content validity explores the degree to which the items of an instrument reflect the construct of a test. Rogers (2010) described content validity as being based on professional judgments of test content relevancy to the content of the test domains, and representation of items to their domains. Therefore analysis of the content is mainly subjectivity of the judges (Allen & Yen, 1979). According to Messick (1990), the judgment results indicate the relevancy of the "test content to the content of a particular behavioural domain of interest" (Messick, 1990, p. 8). However, the question of how one should deal with the results from expert judgment is rarely mentioned in the literature.

2.2.2.4 Methods of Analyzing Content Validity. In a research study by Hellsten (2008), methods of analyzing content validity were identified, and methods on how to analyze judge ratings were also identified. Three major approaches used to analyze judgment ratings were classified: (a) qualitative, (b) descriptive, and (c) quantitative. A complete list of methods under each approach is listed in Appendix C. This study will employ descriptive and quantitative methods due to the significantt time requirement and resources required for qualitative research (e.g., focus groups, and interviews). Three kinds of descriptive analyses (i.e. Item Ambiguity,

Median, and Percentage Agreement) and three quantitative methods (i.e. Content Validity Index, Content Validity Ratio, Content Validity Coefficient) were used for this research study.

CHAPTER 3

3. Development of Item Pool

This chapter focuses on three procedures used for developing and accumulating validity evidence for an instrument. The first procedure of this study involved developing domain definitions, item writing, and the categorization of items. The second procedure involved sending the newly created items to a panel of content experts for an expert judgmental review. The third procedure involved the analysis of the expert ratings using a variety of different statistical methods in order to gather some evidence of content validity. Due to the sequential nature of the procedure, this chapter will include both the methodology employed and the results for the development of the item pool for the proposed Asperger's Syndrome self-screening instrument. The procedures used to develop the item pool will be discussed first followed by a discussion of the criteria for the inclusion of expert judge participants and how the experts were utilized.

3.1 Item Writing and Categorization

3.1.1 Methodology

Items were first gathered from the existing measures of AS for adults. These measures included the RAADS (Ritvo et al., 2008), the AQ (Baron-Cohen et al., 2001), the AAA (Baron-Cohen et al., 2005), the Empathy Quotient (Baron-Cohen & Wheelwright, 2003), and the adult version of the Australian Scale for Asperger Syndrome (Meyer, 2000). Items on the Empathy Quotient (Baron-Cohen & Wheelwright) were also included in this study due to its association with the AAA (Baron-Cohen et al., 2005) and because people with AS often report lower levels of empathy in comparison to the normal population (Baron-Cohen et al., 2005; Lawson, Baron-Cohen, & Wheelwright, 2004). All the items from each scale were compared to each other as

well as to the stated domains to which they were originally developed. This process acted as a check for item redundancy in the final pool of items. If two or more items with similar meaning or wordings were found, the item with the best quality in terms of wording, ease of interpretation, and representation of content was selected. All the selected items were then rewritten and modified to reflect the first person and the context of a self-screening instrument. Items then proceeded for reviewing and categorizing by relevant themes in order to create the new domains. Items that did not appear to fit any of the domains were deleted. Further item screening included examining their content to determine the appropriateness of language level. For example, phrases that required participants to understand high levels of the English language were removed. Furthermore, items that did not match the self-screening context were also removed. For example, items with content that potential self-screeners would not be able to recall were deleted (such as age when speech was first acquired). Where possible, quality of remaining items was improved using the item-writing guidelines and rules developed by Frey et al. (2005) (see Appendix B). Suggestions for item writing provided by DeVellis (2003) and Rogers (2010) were also used as guidelines. Content of the items was then re-examined and categorized according to the domains described by the DSM-IV-TR (APA, 2000). The domains will be discussed further in the following section. The number of items written under each domain were more than required by the final stage, as DeVellis (2003) suggested that having an abundance of items is a safeguard to inadequate internal consistency and construct underrepresentation. A large number of initial items also allows for further item deletion if necessary. It was estimated that about 10 items for each subscale would be included on the final version of the AS self-screen. For content validity purposes of this study, it was proposed that

about four domains would be created and 80 items (20 items for each domain) were to be written.

3.1.2 Creation of Domains

As reviewed in Chapter 2, the diagnostic criteria of AS under the DSM-IV-TR (APA, 2000) include: (a) Qualitative impairment in social interaction; (b) Restricted repetitive and stereotyped patterns of behaviours, interests, and activities; (c) Clinically significant impairment in social, occupational, or other important areas of functioning; (d) No clinically significant general delay in language (e.g., single words used by age 2 years, communicative phrases by age 3 years); (e) No clinically significant delay in cognitive development or in the development of age-appropriate self-help skills, adaptive behaviour (other than in social interaction), and curiosity about the environment in childhood; and (f) Criteria are not met for another specific Pervasive Developmental Disorder or Schizophrenia (APA, 2000). Due to the nature of a self-screening instrument, only the observable criteria of AS were included as domains of this study, therefore only (a) and (b) of the above criteria were included as domains for this new self-screening instrument. Although criterion (c) was also observable, it was not included because one cannot easily determine what contributes *clinically significant* without clinical measures. Up to this point, there were only two domains at the primary stage, underrepresentation of AS was possible, therefore in an attempt to increase representation of AS, more literature reviews were conducted to form additional domains for AS.

3.1.3 Domain Definition Results

Three domains were initially created for this study (i.e. Deficit in Social and Communication Skills**;** Sensorimotor and Setereotyped Patterns of Behaviours; and Stereotyped

and Restricted Repetitive Patterns of Interests and Activities). Upon reviewing the relevant literature further, four categories (i.e. Deficit in Social Communication Skills; Sensorimotor and Setereotyped Patterns of Behaviours; Stereotyped and Restricted Repetitive Patterns of Interest and Activities; and Empathy) were then developed. Two subcategories (i.e. Problems in Pragmatic Language, and Other Problems in Communication Skills) were established from Deficit in Social Communication Skills (to address the qualitative impairment in social interaction described in the DSM-IV-TR; APA, 2000).

Communication skill is a factor that includes many components. It often takes more than one element in order for communication to happen, especially social communication which involves people-to-people reciprocal interactions. Of the many language components, pragmatic language plays a significant role in social communication. The American Speech-Language Association (2009) explained social language use (or pragmatics) as involving three major communication skills: (a) using language for different reasons such as greeting, informing and promising; (b) the change of language based on situation and the listener (for example, talking in a different tone with a baby in comparison to an adult); and (c) following conversation and storytelling rules such as turn-taking during a conversation, and using verbal and non-verbal signals. Individuals with pragmatic problems may display characteristics such as saying inappropriate things at inappropriate times, misuse of eye contacts, and difficulty staying on topic in a conversation. Individuals with ASD often display a problem with pragmatic language use (Martin & McDonald, 2003). These problems can also be traced back to the parents of ASD individuals, in which these parents tend to display more abnormal pragmatic behaviours than control parents (Landa et al., 1992).

After screening the items, due to an insufficient number of items to further divide communication skills into different sub-domains, the sub-domain "Other Problems in Communication Skills" was added to include items that do not fit into pragmatic language, but also relate to communication skills that may be important to distinguish people with AS from the normal population.

Although empathy was not listed as a criterion on the DSM-IV-TR (APA, 2000), studies have demonstrated that individuals with AS and ASD tend to display lower levels of empathy in comparison to control individuals (Baron-Cohen et al., 2005; Baron-Cohen & Wheelwright, 2004). Empathy was originally invented by Titchener (1909) using the German word "einfuhlung" which means "to project yourself into what you observe" (Baron-Cohen & Wheelwright, 2004, p. 163). There were two approaches to empathy research: the affective approach, and the cognitive approach. The affective approach, described by Baron-Cohen and Wheelwright (2004) defines empathy as "an observer's emotional response to the affective state of another" (p. 164), in which the observer's emotion is due to observation of someone else's emotion in order to be considered as empathy. The cognitive approach, defined by Kohler (1929) indicates that empathy needs to be involving the understanding of other's feelings. However, Baron-Cohen and Wheelwright (2004) suggested that it is necessary to include both approaches in understanding empathy as a whole. They also compared empathy with sympathy in which sympathy occurs when the observer actually has the desire to do take action to decrease another person's suffering, after having an emotional response to the other person's distress. Whether the observer would actually take action does not matter, what matters in sympathy is the desire itself to do something. They described Asperger's Syndrome (AS) and Higher Function Autism (HFA) as an empathy disorder, due to these individuals' problems with displaying empathetic

behaviours (Baron-Cohen & Wheelwright, 2004). Therefore, in an effort to be inclusive and avoid construct under-representation, the researcher made the decision to include empathy as a domain for screening AS individuals in this study.

The following are the definitions of the domains employed in this study. Each of the domain definitions were created based on the literature review, and descriptions that appeared most frequently across the literature were chosen as part of the definition. Definitions for the domains *Stereotyped and Restricted Repetitive Patterns of Interests and Activities*, and *Sensorimotor and Stereotyped Patterns of Behaviours* were mainly based on the diagnostic outline provided on the DSM-IV-TR (APA, 2000).

Empathy

The ability to understand and put oneself into others' feelings (Baron-Cohen & Wheelwright, 2004). It is one's reaction upon observing others' experiences (Davis, 1980).

Stereotyped and Restricted Repetitive Patterns of Interests and Activities

Refers to an intense interest on certain things and/or activities. This also includes strictly following a routine by an individual (APA, 2000).

Sensorimotor and Setereotyped Patterns of Behaviours

Defined as displaying of unusual motor actions by an individual regularly (APA, 2000). Hyper- and/or hypo- sensitivity on objects/things (Barnhill, 2007; Hurlbutt & Chalmers, 2004) are also included in this category.

Deficit in Social Communication Skills: Problems in Pragmatic Language

Refers to difficulties in social use of language -understanding and applying the rules in social communication. This involves the lack of abilities with respect to: use of language, change of language, and following rules of communication (American Speech-Language-Hearing

Association, 2009). Examples include the use of facial expressions, verbal and non-verbal signals.

Deficit in Social Communication Skills: Other Problems in Communication Skills

Include the displaying of inappropriate motor actions by an individual during social interactions, personal feelings of inability to connect with others, inability to understand strategies required in a social situation, and negative preference on reacting to social situations.

3.1.4 Item Writing and Categorization Results

After combining all five of the existing measures, there were a total of 288 items. Next, the items were reviewed for the presence of double-barreled items or items containing more than one meaning. After removing the double-barreled phrases from the long items by separating them into more than one item, 318 items existed (see Appendix D). Of the 318 items, 50 items were removed due to repetition of content and overlap with other items, 31 items were removed due to their need for a high language level on the part of the respondent, or they did not fit into the content of a self-screening instrument, and 69 items were removed because they did not fit into the domains established for this study. Item categorizations were done by matching the definitions of domains with the content of items, and each item were put into only one domain that matched the best.

3.2 Preliminary Item Screening and Judgment

3.2.1 Methodology

Due to the fact that the author performed all the processes of rewriting and re-categorizing the items into their new domains, potential error or bias may occur. To prevent the existence of bias and to ensure the quality of items before they were sent out to expert judges,

revisions of the domain definitions and items were first performed by a group of *pre-judges* consisting of a Psychometrician, a Speech and Language Pathologist, and a class of students from a graduate level Advanced Test Theories and Instrument Construction course at the University of Saskatchewan. The group of students was chosen due to their current knowledge and experience with item development rules. The *pre-judges* were asked to examine the quality of items (i.e. choice of words, relevancy to domains, grammatical and spelling of items), they were also asked to review the domain definitions as a quality check. The preliminary process of item screening also served the purpose of removal of items that demonstrated misfit, were overlapped, and/or were questionable.

3.2.2 Results

Eighty-two items were deleted upon the completion of pre-screening by the *pre-judges*. These judges also performed revisions of the items including re-categorization of items, grammatical, spelling, and phrasing corrections, screening for redundant items, and judgment on the importance of each item under their domains. The final list of items can be found on Appendix E under each Item Content Review Rating Form. There are 16 items classified under the domain Empathy, 16 items were categorized under Sensorimotor and Stereotyped Patterns of Behaviours, 18 items were classified under Stereotyped and Restricted Repetitive Patterns of Interests and Activities, 18 items were categorized under Problems in Pragmatic Language, and 18 items were classified under Other Problems in Communication Skills.

3.3 Item Content Rating Review Forms

Items were arranged into domains on the Item Content Rating Review Forms. A detailed description/definition of the category was provided on the top of the form so participants could

refer to it during the rating process. Items believed to fit each of the domains were listed on the left, and a rating scale of 0 (No Fit) to 4 (Excellent Fit) on the right side. Additional space was also provided for judges to write their suggestions on how items could be better revised. Room was also provided for further comments on the bottom of the form. One *lie* or marker item was included in each subscale in order to help identify aberrant judges. These *lie* items were written by the researcher and were intended to be irrelevant to the domain. If an expert judge has carefully examined each item on the Item Content Rating Review Form, then they should correctly identify these *lie* items as being irrelevant to the domain. The Item Content Rating Review Form can be found in Appendix E.

The associated *lie* items for each of these subscales were item numbers 10, 3, 14, 17, and 5 respectively for the domains Empathy, Sensorimotor and Stereotyped Patterns of Behaviours, Stereotyped and Restricted Repetitive Patterns of Interests and Activities, Problems in Pragmatic Language, and Other Problems in Communication Skills. Excluding the *lie* items, there were a total of 86 items that were reviewed by the expert panel members.

3.4 Expert Judgment

3.4.1 Methodology

3.4.1.1 Participants. It was proposed that the expert judgments would be conducted by at least five judges who are professionals in ASD, in either the fields of Speech-Language Pathology or School Psychology and at least five experts who are familiar with the process of instrument development or were doctoral trained or training to be Psychometricians. At least five judges in each of those fields would ensure there were at least 10 experts panel members. However, more judges are always preferred as there may be potential aberrant judges that require

removal. Psychometricians were included in this study due to the psychometric component of this study, and to greater ensure the overall quality of the items. Speech-Language Pathology judges were also included because of the presence of the pragmatic language domain in this study and to ensure the definition and their associated items were accurately categorized. Experts in School Psychology were chosen due to their knowledge and experience with individuals with AS in a school environment.

Recruitment of participants was done by conducting an Internet search of Canadian university faculty. Potential participants who met the criteria discussed in the next paragraph were invited to participate through electronic mail. Personal connections of the researcher who is in the field of psychometrics were also used. Experts from the field of ASD were recruited with the help of Dr. Janine Montgomery at the University of Manitoba, due to her expertise in the field.

In order to be qualified to participate, the Autism Spectrum Disorder experts needed to hold a minimum of a Master's degree in Special Education or Psychology, they were required to work in the field of Autism for at least 5 years, and they must have had experience with at least 20 individuals with Asperger's Syndrome. The researcher set the requirement for five years of experience because it is unlikely that a practitioner with five years of experience would be considered a novice in the field. The researcher has also estimated that if an expert was in the field for a minimum of five years, he/she needs to work with an average of at least four people of AS each year in order to be familiar with individuals with AS. People with AS and ASD are often portrayed as unique individuals as they as very different from one another, therefore, an estimation of 20 individuals appears to be a reasonable number. Experts in psychometrics were required to have at least a Master's degree in the relevant field and have taken graduate level

courses on instrument development, they were also required to be trained as a Psychometrician or undergoing training in that field. School Psychologists and Speech and Language Pathologists were required to have a Master's degree in their respective field and five years of experience. This education level was set because a Master's degree is usually a minimum education requirement in order for an individual to enter the field.

3.4.1.2 Expert Judgment Procedures. People who agreed to participate were sent a package containing: (a) the informed consent form; (b) the research introduction letter explaining the nature of the study; (c) a participant information form to ensure the qualification of the judges; (d) an instruction sheet with domain definitions; (e) the Item Content Rating Review Forms; and (f) a self-addressed stamped envelope. Participants were given 14 days to complete the package and mail it back to the researcher.

Participants' were informed that their responses would always be kept confidential and that only the researcher and her supervisor would have access to the data. Participation was always voluntary and individuals were free to withdraw anytime. At any time, if a participant should wish to withdraw from the study, they were instructed to not return the envelope. Envelopes that were not received 10 business days after the 14-day period were considered to be a withdrawal.

3.4.2 Results

3.4.2.1 Participants. Following the internet search for expert judges form universities across Canada and receiving a list of potential judges from Dr. Montgomery, 21 invitations to participate were sent via electronic mail. Of the 21 individuals, 12 individuals agreed to participate and the packages were mailed out to these experts. A final total of nine packages were

received, resulting in a response rate of 75% of those packages that were mailed out. Of the nine participants, five judges belonged to the field of Psychometrics; one judge belonged to the ASD and the Psychometric field; one judge was from the field of ASD and Speech-Language Pathology, one judge belonged to the field of ASD and School Psychology, and one judge came from the field of School Psychology. Of the experts in the ASD field, at least 10 years of applied experience were obtained by these individuals, and they have each worked with a minimum of 40 people with AS. Experts in Psychometrics were all trained as a Psychometrician and had a minimum of a Master's degree. The Speech-Language Pathology and School Psychology experts all had a minimum of 5 years of applied experience in their field. Of the judges participating, eight judges were from Canada, and one judge was from the United States of America. The following table listed the distribution of judges:

Table 3.1 Distribution of Judges' Expertise

Expertise	Number of Judges
Psychometrics	5
ASD + Psychometrics	1
ASD + Speech-Language Pathology	1
ASD + School Psychology	1
ASD	1

Although there were only nine expert judges in total, the distribution of expertise across judges seemed to be reasonable. The emphasis on the psychometric quality of this measure is represented by the number of Psychometric judges included in this study. One potential problem was the small number of experts in ASD and the Speech-Language field. There were a total of three ASD judges in comparison to the expected five; however, one "preliminary" judge was in the field of both ASD and Speech-Language Pathology.

CHAPTER 4

4. Content Validity – Analysis of Ratings

This chapter discusses the second stage of this research – analyzing the item ratings conducted by the expert panel members. The methodology and the results of identification of any potential aberrant judges will first be discussed. This discussion was then followed by an examination of the procedures used to examine and analyze the judges' ratings using both descriptive and quantitative analyses. Results of these analyses were used to help decide which items should be kept and which items should be excluded from the future item pool. Lastly, the analytical methods used to compare judges' ratings were examined in an attempt to distinguish the best method(s) to be used in future research.

4.1 Methodology

As presented in Chapter 3, nine judges participated in this research. These experts were in the field of Psychometrics (n = 5); ASD (n = 1); ASD and Psychometric (n = 1); ASD and Speech-Language Pathology (n = 1); and ASD and School Psychology (n = 1). All experts except for the ASD and Psychometric expert were from Canada.

Upon receipt of the expert panel members' ratings, and separating the participants' names from the data, the ratings of the judges were entered into the Microsoft Excel computer program. In order to accumulate evidence of content validity, the expert panel members' ratings for each of the items on each of the subscales were compared and contrasted. As discussed in Chapter 2, content validity is based on professional judgments of test content relevancy to the content of the test domains, and representation of items to their domains (Rogers, 2010). Three major approaches to analyze judgment ratings were classified by Hellsten (2008): (a) qualitative, (b)

descriptive, and (c) quantitative. The associated methods for each approach are listed in Appendix C. Due to the time and resource restraints in this research, qualitative analyses (e.g., interviews, open-ended feedback, and focus groups) were not employed. In this research, three kinds of descriptive analyses (i.e. Item Ambiguity, Median, and Percentage Agreement) and three quantitative methods (i.e. Content Validity Index, Content Validity Ratio, Content Validity Coefficient) were used to determine the quality of the items. These methods will be discussed below in their associated sections. The methods were then further compared to determine their overall agreement, and to determine which method(s) is/are the best to use for content validity analysis.

4.1.1 Identifying Aberrant Judges

Although the judges were experts in their selected field, there is the possibility that some judges scored the items in an aberrant manner. For example, lack of understanding of the procedure or directions, inattention during the ratings, lack of time, and/or personal motivation may result in measurement error across the ratings. In this study, two methods were used to identify aberrant judges. The first method examined if each judge was able to correctly identify the "lie item" placed within each domain while the second method examined the degree of inter-judge agreement. Inter-judge agreement helps to identify which expert(s) are discrepant raters as compared to the rest of the judges. If necessary, following the identification of any aberrant judges, decisions were made to remove such aberrant judges from further analysis (i.e., descriptive and quantitative analysis).

4.1.1.1 Detecting the lie items. A table was created which lists all the domains and the number of *lie* items accurately detected by each judges. This process allowed the researcher to identify which judge, if any, should be excluded due to his/her potential inaccuracy of ratings

(Hellsten, 2009a). As all items were rated on a scale ranging from 0 No Fit to 4 Excellent Fit, a *lie* item should have a low rating (i.e., either 0 or 1) because it was specifically designed to not fit the domain. If the judges read the domain definition and examined each item carefully, they should be able to correctly identify the *lie* item by rating it low. Each judge who correctly identifies the *lie* item was identified by a check mark. The percentage of lie items correctly identified was also calculated for each judge. There are a total of five *lie* items (i.e. one in each subscale), and the researcher set the criteria that three or more *lie* items (60%) correctly identified would be considered as acceptable. This criteria was based on the rational that if the judges correctly identified 50% or more of the *lie* items, then it is more likely that the correct response was not made due to chance.

4.1.1.2 Inter-judge agreement (Judges' Discrepancy From the Median - JDM). A second method used to help identify aberrant judges was to identify the discrepancy between each judge's rating score from the median rating provided by all judges on each item. The discrepancy was then summed across all subscales/domains for each judge (JMD_{js}; Hellsten, 2009a; Rogers, 2010). The formula to represent the inter-judge agreement is represented as:

$$JDM_j = \sum_{k=1}^{K} |X_{kj} - Md_k| \quad \text{(Rogers, 2010)} \qquad (4.1)$$

Where X_{Kj} is Judge j's rating on subscale K, and Md_K is the median of item K. JDM of each Judge j is equal to the sum across his/her ratings on each item minus the median of that item. Expert(s) with exceptionally high JDM score(s) in comparison to other judges were compared with their results from the identification of *lie* items. If both methods indicate the judge did not perform as desired (i.e., missed identifying the lie items or scored items

substantially different from the other judges), his/her ratings were eliminated from further analyses of descriptive and quantitative analyses of items.

4.2 Results

Results of identifying aberrant judges by using the method of *lie* item accuracy and the calculation of JDM are presented below. The decision as to which judges were ultimately removed will follow.

4.2.1 Identifying Aberrant Judges

4.2.1.1 Detecting the lie *items*. A table representing the accuracy in identifying the *lie* items by judges is shown in Table 4.1. Short terms of the subscale names have been used: EMP for Empathy; SIA for Stereotyped and Restricted Repetitive Patterns of Interests and Activities; SPB for Sensorimotor and Stereotyped Patterns of Behaviours; DPL for Deficit in Social Communication Skills: Problems in Pragmatic Language; and DCS for Deficit in Social Communication Skills: Other Problems in Communication Skills. Across all five subscales, the number of judges who correctly identified the *lie* item ranged from 0 to 5 with a median of 4. Three of the judges identified all five *lie* items (Judge 2, 3, 4), and one judge did not identify any of the five *lie* item (Judge 7). It was determined by the researcher, that a minimum of 60% accuracy rate (3 out of 5 *lie* items) should be obtained for an expert to be maintained in the study. Thus, judges 5, 7, and 8 are potential aberrant judges, with 0%, 20%, and 20% accuracy rate in identifying the *lie* items respectively. However, due to the low number of judges in this study, before any decision to remove judges was made, the results of the second method of identifying potentially aberrant judges was examined.

Table 4.1 *Identification of Lie Items by Judges*

Judge	***Subscale***					***# ID***	***% ID***
	EMP	*SIA*	*SPB*	*DPL*	*DCS*		
Judge 1	✔	✔	-	✔	✔	4	80%
Judge 2	✔	✔	✔	✔	✔	5	100%
Judge 3	✔	✔	✔	✔	✔	5	100%
Judge 4	✔	✔	✔	✔	✔	5	100%
Judge 5	-	-	-	-	-	0	0%
Judge 6	✔	✔	✔	-	-	3	60%
Judge 7	-	-	-	✔	-	1	20%
Judge 8	✔	-	-	-	-	1	20%
Judge 9	✔	✔	-	✔	✔	4	80%
Total # of Judges ID	7	6	4	6	5		
Overall % of Judges ID	77.78%	66.67%	44.44%	66.67%	55.56%		

Short Terms of Subscale
EMP – Empathy
SIA – Stereotyped and Restricted Repetitive Patterns of Interests and Activities
SPB – Sensorimotor and Stereotyped Patterns of Behaviours
DPL – Deficit in Social Communication Skills: Problems in Pragmatic Language
DCS – Deficit in Social Communication Skills: Other Problems in Communication Skills

4.2.1.2 Inter-judge agreement (JDM). The results of the ratings of inter-judge agreement are shown in Table 4.2. Due to the fact that some judges did not complete a rating for some items, by being more conservative, all possible deviation scores for any un-attempted items were added into each judge's total JDM. They were treated as ratings that were fully deviated from the median of that item. Results show that Judge 1 had an exceptionally high JDM score (JDM_1 = 108) as compared to the others; therefore Judge 1 may be a potential problem that could affect subsequent analyses. However, Judge 1 had an 80% accuracy rate in correctly spotting the *lie* item. Table 4.3 below presented the results of judges on both methods for comparison.

Table 4.2 *Inter-Judge Agreement (JDM scores)*

	Subscale JDM of Judges					
Judge	*EMP*	*SIA*	*SPB*	*DPL*	*DCS*	***Judge Total JDM***
1	12	22.5	30.5	21	22	108
2	5	4.5	6.5	9	6	31
3	9	12.5	8.5	10 (1R)	6	46 + 1R (4 scores dev. from DPL item 3) = 50
4	13	13.5	7 (1R)	10	4 (2R)	47.5 + 1R(2.5 scores dev. from SPB item 1) + 2R(3+3 scores dev. from DCS item 3&4) = 56
5	6	7.5	8.5	8	13	43
6	6	4.5	5.5	5	11	32
7	13 (1R)	17 (1R)	12.5	13	14	69.5 + 1R (3 scores dev. from EMP item 3) + 1R (3.5 scores dev. from SIA item 8) = 76
8	9	10.5	12.5	11	10	53
9	15	13.5	12.5	4	8	53
Min	5	4.5	5.5	4	6	
Max	16 (with max deviation of 3 scores from 13 to Median)	22.5	30.5	21	22	

- Each R represents any un-attempted items by the judges. The Rs are automatically assumed to be all possible deviated scores from the median for that particular item in the subscale.

Table 4.3 Comparison of Methods on Identifying Aberrant Judges

	Method	
Judge	*ID of Lie Items*	*Inter-Judge Agreement: Total JDM*
1	80%	108
2	100%	31
3	100%	50
4	100%	56
5	0%	43
6	60%	32
7	20%	76
8	20%	53
9	80%	53

According to the first method, Judge 1 was able to identify all of the *lie* items. However, Judge 1 had the highest JDM of all judges. Before any decision was made, the field of expertise of Judge 1 was considered. Judge 1 was an expert in Psychometric, and one explanation for his/her ability to correctly identify the *lie* items may be due to the individual's ability to spot items that clearly distinguish themselves from others (or maybe the *lie* item were too easily identified). Although Judge 1 did very well on Method 1, his/her high JDM cannot be ignored, because it clearly shows that his/her ratings are very different from other judges. Therefore after considering the expertise of Judge 1, in which he/she may have had limited AS knowledge, Judge 1's ratings were removed to prevent the scores from affecting further analyses (e.g., effects on the median).

Judge 7 had the next highest JDM score. According to Rogers (2010), if one judge has been removed, the median should be recalculated with the rest of the judges and the JDM needs to be recalculated to determine if there is a second aberrant judge. Following this recommendation, the median and JDM were recalculated. Table 4.4 shows the Inter-Judge Agreement with Judge 1 removed. After performing another calculation of the JDM with Judge 1

removed, Judge 7 was still showing a JDM score of at least 10 points higher than the next highest JDM score even when the un-attempted items were not accounted for. Before Judge 7 was removed, Rogers (2010) suggested that the comments of the experts should be considered before the removal of any additional judges, because researchers should attempt to retain as many experts in the panel as possible. A decision to remove Judge 7 was determined by the researcher when the comments read were mainly about wording changes to the items. In addition, Judge 7 was only able to correctly identify 20% of the *lie* items.

Table 4.4 *Inter-Judge Agreement with Judge 1 Excluded*

	Subscale JDM of Judges					
Judge	*EMP*	*SIA*	*SPB*	*DPL*	*DCS*	***Judge Total JDM***
2	6	5.5	8	8.5	7	35
3	9	11.5	8	11.5 (1R)	6	46 + 1R (4 scores dev. from DPL item 3) = 50
4	12	12.5	6 (1R)	8.5	4 (2R)	43 + 1R(3 scores dev. from SPB item 1) + 2R(3+3 scores dev. from DCS item 3&4) = 52
5	6	5.5	8	8.5	13	41
6	7	6.5	6	5.5	11	36
7	12 (1R)	17.5(1R)	13	12.5	13	68 + 1R (3 scores dev. from EMP item 3) + 1R (4 scores dev. from SIA item 8) = 75
8	8	10.5	12	11.5	9	51
9	16	13.5	12	3.5	9	54
Min	5	4.5	5.5	4	6	
Max	16 (with max deviation of 3 scores from 13 to Median)	22.5	30.5	21	22	

4.2.1.3 Removal of experts. Judge 1 and Judge 7 were removed from further analyses after examining the calculations for inter-judge agreement. After reviewing the first method on correctly identifying the *lie* items, the researcher decided to also remove Judge 5 due to his/her inability to identify any of the *lie* items (i.e., 0%). Another factor that led to the removal of these judges was due to their background of expertise. All three of the aberrant judges belonged to the Psychometric field (i.e. Judge 1 and 7 were experts in Psychometrics, and Judge 5 belonged to both Psychometrics and ASD). It is possible that these judges were using a different perspective during their rating process in comparison to other experts, this problem will further be discussed

in discussions. From this point, ratings of Judge 1, 5, and 7 were eliminated from further descriptive and quantitative analyses. The total number of expert panel members for this research was now six expert judges (i.e. three Psychometric experts, one ASD expert, one ASD and Speech-Language Pathology expert, and one ASD and School Psychology expert).

4.3 Descriptive Analysis

4.3.1 Methodology

Descriptive analyses were then conducted on the ratings from the remaining expert panel members. First, the median item rating, or the number that indicates the midpoint of all ratings was calculated for each subscale. A higher median value indicates a more relevant item. Following the work of Hellsten (2008), based on a scale of 0 to 4, an item with a median of 2.75 or above was considered as acceptable in this study. Second, the item ambiguity score or the range (R_K) of scores for each item was calculated. Item ambiguity is calculated as follows:

$$R_K = X_{KjH} - X_{KjL} + 1 \quad (4.2)$$

Where X_{KjH} is the item's highest rating; and X_{KjL} is the lowest rating (Hellsten, 2009a; Rogers, 2010). As the range of possible ratings range from 0 to 4, items with lower values are desired as they indicate consensus among judges. Due to the rating scale of 0 to 4 used in this research, a range of three or more between scores (or R_k of 4 or higher) was considered ambiguous. Rogers (2010) suggested that items with high ambiguity should be seen as "yellow flashing lights". Items with high ambiguity should not be easily removed before other evidence has been collected. Caution should also be paid even when the items show low ambiguity because low ambiguity does not necessarily mean an item is representing a domain well. Judges may all be agreeing that an item does not fit well with lower ratings, which may also lead to a low item

ambiguity score. Therefore, before any item is removed based on this score, judges' ratings of how well the item fits the category should be examined.

Third, percentage agreement was calculated to indicate the percentage of judges who agreed the item was a good fit to the category specified. The question "Is this item essential to the domain" was asked at the end of each item and raters were asked to choose 'Yes' or 'No'. The formula for calculation of percent agreement is:

$$\text{Percentage agreement} = \left(\frac{\text{\# of judges rated "yes"}}{\text{Total \# of judges}}\right) \text{X } 100 \tag{4.3}$$

4.3.2 Results

Descriptive analyses were used to show the properties of the items. The median for each item was calculated as a measure of the central tendency and can be found in Table 4.5. A median of 3 or above means at least 50% of the judges gave an item a rating of 3 or 4. Generally, the items received quite high ratings.

Table 4.5 *Median of Items by Subscale*

	Subscale				
Item	*EMP*	*SIA*	*SPB*	*DPL*	*DCS*
1	**4**	**3**	2	**3**	**3**
2	2	1.5	**4**	**3**	**3**
3	**3**	**3**	1	**4**	**3**
4	**3.5**	2.5	2.5	**3.5**	**3**
5	**3**	2	**3**	**3**	LIE
6	**3**	**3**	**3**	**3**	**3**
7	**3.5**	**3.5**	**3**	**3.5**	2
8	**3.5**	**3.5**	**4**	**4**	**3.5**
9	**3**	**4**	**3**	**3**	**3**
10	LIE	**4**	**3**	**4**	**3**
11	**3.5**	2.5	2	**3.5**	**3**
12	**3**	**4**	**3.5**	**3**	**4**
13	**4**	**3**	**3**	**4**	**4**
14	**3**	LIE	**3.5**	**3.5**	**4**
15	**3**	**4**	**3**	**3**	**3.5**
16	2	2	**4**	**3**	**4**
17	**3**	**3**	**3.5**	LIE	**3**
18	-	**3**	-	**3**	**3**
19	-	**4**	-	**3**	**4**

Note. Bolded number indicates acceptable median

4.3.2.1 Item ambiguity. Item ambiguities were calculated using Microsoft Excel computer software and the results are shown in Table 4.6. Recall that item ambiguity (R_k) was calculated by using the highest rating minus the lowest rating for each item plus one. Items with high ambiguity are unacceptable, as they indicate high discrepancy among judges' perception on their fitting in domains. The items with high ambiguity (i.e. item 2, 3, 5, 15, 16 for EMP; item 2, 4 for SIA; item 10, 11, 17 for SPB; and item 1 for DCS) should be treated with caution as they may not fit the domain as well and may be earmarked for future deletion. Recall that Rogers (2010) suggested that items with high ambiguity should be seen as "yellow flashing lights" and therefore these items should not be deleted at this point.

Table 4.6 *Item Ambiguity*

	Subscale				
Item	*EMP*	*SIA*	*SPB*	*DPL*	*DCS*
1	1	3	3	2	**4**
2	**5**	**5**	2	2	3
3	**5**	3	LIE	2	3
4	3	**4**	3	2	2
5	**4**	2	2	2	LIE
6	3	3	2	3	2
7	3	2	3	2	3
8	2	2	3	2	2
9	3	2	1	3	2
10	LIE	1	**4**	2	2
11	3	3	**5**	2	1
12	3	1	3	3	2
13	2	3	2	2	2
14	3	LIE	2	3	2
15	**4**	3	2	2	2
16	**4**	3	2	3	2
17	2	3	**4**	LIE	3
18	-	3	-	2	3
19	-	2	-	2	2
Number of ambiguous items	5	2	3	0	1

Note. Bolded numbers: Range set at 3 or more is problematic (an item ambiguity score of 4 or 5)

4.3.2.2 Percentage agreement. Percentage agreement was also calculated for each item. Some of the judges did not answer certain question of "Yes" or "No" and thus percentage agreement calculations only included the number of people who responded. As seen in Table 4.7, the majority of the items were rated as essential to the domain by over 50% of the judges. However, like item ambiguity, the items with low percentage agreement should be treated with caution instead of deletion at this point. Due to the low number of judges in this study, percentage agreement can easily fluctuate, and therefore decisions regarding the deletion of items should not be made on the basis of only one calculation. For comparison of methods in

deciding which items to keep, a percentage agreement of 80% was set as high agreement between judges, this value was determined when a total of six judges were considered, in which the 80% represents almost all judges in the panel except for one. It also applies to the case of missing items where only scores of five judges were used (i.e. 4 out of 5 judges = 80%).

Table 4.7 *Percentage Agreement*

	Subscale				
Item	*EMP*	*SIA*	*SPB*	*DPL (only 5 judges responded)*	*DCS (only 5 judges responded)*
1	100%	100%	40% (5 responses)	80%	75% (4 responses)
2	16.67%	16.67%	100%	100%	80%
3	66.67%	50%	LIE	100% (4 responses)	80%
4	83.33%	40% (5 responses)	33.33%	100%	60%
5	66.67%	33.33%	83.33%	60%	LIE
6	50%	50%	100%	80%	100%
7	66.67%	66.67%	66.67%	100%	75% (4 responses)
8	100%	100%	83.33%	100%	100%
9	50%	100%	83.33%	100%	80%
10	LIE	100% (5 responses)	50%	100%	100%
11	66.67%	50%	33.33%	100%	80%
12	66.67%	100%	60% (5 responses)	100% (4 responses)	80%
13	100%	100% (5 responses)	100%	100%	100% (6 responses)
14	83.33%	LIE	100%	60%	100% (6 responses)
15	33.33%	66.67%	83.33%	80%	100% (6 responses)
16	33.33%	16.67%	100%	80%	100% (6 responses)
17	83.33%	50%	66.67%	LIE	83.33% (6 responses)
18	-	66.67%	-	75% (4 responses)	83.55% (6 responses)
19	-	100%	-	100%	100% (6 responses)

4.4 Quantitative Analysis

4.4.1 Methodology

4.4.1.1 Content validity index. The content validity index for each individual item (CVI; Lynn, 1986) is the percentage of judges that rated the item as 3 or 4 (based on the rating scale of 0 to 4 where 4 represents excellent fit). The formula is represented as:

$$\text{CVI} = \left(\frac{\text{\# of judges rated 3 or 4}}{\text{Total \# of judges}}\right) \text{X } 100 \tag{4.4}$$

The CVI is expressed as a percentage. To determine what an acceptable CVI is, Polit, Beck, and Owen (2007) suggested the item *CVI* values should be 1.00 for expert panels of three or four judges, 0.80 for panels of 5 members, and 0.78 for larger expert panels. An acceptable CVI in this study was determined to be 0.80 (i.e. 80%) or above since there were six judges in this study.

4.4.1.2 Content validity ratio. The content validity ratio (CVR; Lawshe, 1975) is calculated using the following formula:

$$\text{CVR}_i = \frac{\left[n_e - \left(\frac{N}{2}\right)\right]}{\left(\frac{N}{2}\right)} \tag{4.5}$$

CVR_i is the value of CVR for the i^{th} item, n_e is the number of experts indicating that the item is *essential*, and N is the number of experts on the panel. CVR values range from -1 to +1. Negative values indicate that less than half of the experts rated the item as *essential* whereas positive values indicate that more than half of the experts rated the item as *essential*, and the number will equal to zero when exactly half of the judges rated the item as *essential* (Hellsten, 2008; Lawshe, 1975). When using the CVR, usually two assumptions are made (Hellsten, 2008;

Lawshe, 1971). The first assumption is that when more than half of the judges rate an item as *essential*, then the item must hold at least some evidence of content validity. The second assumption states that the greater number of judges (when there are more than half of the judges) rate the item as *essential*, the item has higher evidence of content validity. For the purpose of calculating CVR, the question "Is the item essential to the domain" was added to the Item Content Rating Review Forms for judges to circle their responses (i.e. "Yes" or "No").

4.4.1.3 Content validity coefficient. The content validity coefficient (VI_K; Aiken, 1985) is calculated with the formula:

$$VI_k = S/[j(c\text{-}1)] \quad (4.6)$$

where S is the sum of S_j, ($S_j = r_j\text{-lo}$); r_j is the rater j's rating, and 'lo' is the lowest validity category. The j in the VI_k formula is the total number of judges, and c is the number of rating categories (five in this case; from 0-4). The closer the coefficient is to 1, the higher content validity an item has. To determine whether the calculated coefficient is significant, a comparison is made with a table of the *Right-Tail Probabilities (p) for Selected Values of the Validity Coefficient (V)* by Aiken (1985, p. 134). This table, according to Aiken (1985) supplies the *p* and *V* values that only have "right-tailed probabilities close to but not greater than the 0.01 and 0.05 levels" (Aiken, 1985, p. 133).

After all the calculations were made, all the results for each item were then summarized within their respective subscales. A check mark will be used to indicate when the ratings for the item meet the criteria set for each method. Items that are shown to have little agreement between methods will be discarded. A final table of specification indicating the number of items belonging to each subscale will then be created.

4.4.2 Quantitative Analysis Results

4.4.2.1 *Content validity index*. The CVI for each item was calculated and is shown in Table 4.8. Numbers marked with an "*" were significant. In this study, similar to percentage agreement, a CVI of 80% or above was considered as acceptable because with six judges, 80% represents almost all judges in the panel except for one. It also applies to the case of missing items where only scores of five judges were considered (i.e. 4 out of 5 judges = 80%). This number also meets the recommendation suggested by Polit et al. (2007). Results of the CVI for items identified eight items in the EMP subscale; 10 items from the SIA subscale; 10 items on the SPB subscale; 15 items on the DPL subscale; and 15 items on the DCS subscale that contain evidence of content validity.

Table 4.8 CVI Results

	Subscale				
Item	*EMP*	*SIA*	*SPB*	*DPL*	*DCS*
1	100.00*	83.33*	40.00	100.00*	66.67
2	16.67	16.67	100.00*	100.00*	83.33*
3	83.33*	66.67	LIE	100.00*	80.00*
4	83.33*	50.00	50.00	100.00*	60.00
5	66.67	33.33	83.33*	66.67	LIE
6	66.67	66.67	100.00*	83.33*	100.00*
7	66.67	100.00*	83.33*	100.00*	33.33
8	100.00*	100.00*	83.33*	100.00*	100.00*
9	66.67	100.00*	100.00*	100.00*	100.00*
10	LIE	100.00*	66.67	100.00*	100.00*
11	83.33*	50.00	33.33	100.00*	100.00*
12	66.67	100.00*	66.67	100.00*	100.00*
13	100.00*	83.33*	100.00*	100.00*	100.00*
14	83.33*	LIE	100.00*	66.67	100.00*
15	66.67	83.33*	100.00*	100.00*	100.00*
16	16.67	33.33	100.00*	6.67	100.00*
17	100.00*	83.33*	66.67	LIE	83.33*
18	-	66.67	-	100.00*	83.33*
19	-	100.00*	-	100.00*	100.00*
# of significant items	8	10	10	15	15

4.4.2.2 Content validity ratio. The table provided by Lawshe (1975, p.568) was used to identify the minimum CVR for each item. The minimum CVR for each item to be considered as acceptable was .99 for a one-tailed test at the 95% confidence level. The CVR results can be found in Table 4.9 below. Numbers marked with an "*" were acceptable. Using the criteria stated above, three items were identified as providing evidence of content validity in EMP subscale, seven items from the SIA subscale, five items from the SPB subscale, 11 items on the DPL, and eight items on the DCS subscale. Due to the more narrow restriction on the acceptable CVR, fewer items have been flagged as being relevant and representative with the CVR method.

Table 4.9 *CVR Results*

	Subscale				
Item	*EMP*	*SIA*	*SPB*	*DPL*	*DCS*
1	1*	1*	-0.20	0.60	0.50
2	-0.67	-0.67	1*	1*	0.60
3	0.33	0	LIE	1*	0.60
4	0.67	-0.20	-0.33	1*	0.20
5	0.33	-0.33	0.67	0.20	LIE
6	0	0	1*	0.60	1*
7	0.33	0.33	0.33	1*	0.50
8	1*	1*	0.67	1*	1*
9	0	1*	0.67	1*	0.60
10	LIE	1*	0	1*	1*
11	0.33	0	-0.33	1*	0.60
12	0.33	1*	0.20	1*	0.60
13	1*	1*	1*	1*	1*
14	0.67	LIE	1*	0.20	1*
15	-0.33	0.33	0.67	0.60	1*
16	-0.33	-0.67	1*	0.60	1*
17	0.67	0	0.33	LIE	0.67
18	-	0.33	-	0.50	0.67
19	-	1*	-	1*	1*
# of acceptable items	3	7	5	11	8

4.4.2.3 *Content validity coefficient*. The VI_K value for each item is shown in the Table 4.10 below:

Table 4.10 *VI_K for Each Item*

	Subscale				
Item	*EMP*	*SIA*	*SPB*	*DPL*	*DCS*
1	1.00*	.79*	.65	.83*	.67
2	.46	.38	.92*	.83*	.79*
3	.67	.71	LIE	.90*	.75
4	.83*	.63	.67	.88*	.65
5	.71	.58	.71	.67	LIE
6	.71	.75	.79*	.79*	.79*
7	.79*	.88*	.75	.88*	.63
8	.88*	.88*	.92*	.92*	.88*
9	.71	.92*	.75	.79*	.83*
10	LIE	1.00*	.67	.92*	.79*
11	.83*	.71	.50	.88*	.75
12	.71	1.00*	.79*	.75	.92*
13	.92*	.75	.83*	.92*	.96*
14	.79*	LIE	.88*	.79*	.92*
15	.63	.88*	.79*	.79*	.88*
16	.42	.63	.92*	.75	.92*
17	.79*	.79*	.75	LIE	.79*
18	-	.71	-	.83*	.79*
19	-	.92*	-	.83*	.92*
# of significant items	8	9	8	15	13

Three values were used in this study due to the difference in number of judges for missing items. These significant values were V= .79, *p*= .029 for six raters, V= .80, *p*= .040 for five raters, and V= .88, *p*= .024 for four raters. The VI_K for each item were compared using these significant V values and an "*" has been placed beside each significant VI_K. There were eight significant items on the EMP subscale, nine items on the SIA subscale, eight items on the SPB, 15 items on the DPL, and 13 items on the DCS subscale.

4.4.2.4 Agreement across all methods. Items that have fully satisfied all criteria were identified with a check mark in Table 4.11. There were very few items that met this standard.

Only three items on the EMP, six items on the SIA, five items on the SPB, ten items on the DPL, and eight items on the DCS subscale fully satisfied all criteria. Comparison tables for each domain have been created showing the contrast of each method in Table 4.11. The judges' comments for the low agreement items were also included. The comments for high agreement items are mainly for suggestions on wording. A discussion on ranking these methods will be found in Chapter 5.

Table 4.11

Table 4.11.1 *All Ratings for Empathy*

	Method						Comments
Item	*Median*	*Item Ambiguity*	*% Agreement*	*CVI*	*CVR*	VI_K	
1	✔	✔	✔	✔	✔	✔	
2							"Pertains more to self than of others' feelings" "Statement is too broad; what is meant by 'experiences'"
3	✔			✔		✔	"Not the same as empathy"
4	✔	✔	✔	✔		✔	
5	✔						"Perhaps avoid idioms, as this may be confusing to some individuals with AS" "Use of always"
6	✔	✔					"Need to be reworded for clarity by removing negative language"
7	✔	✔				✔	"Need to be reworded for clarity by removing negative language" "Feelings and response are different actions" "Maybe respond appropriately" instead of simply 'respond'"
8	✔	✔	✔	✔	✔	✔	
9	✔	✔					"May need to be more specific about 'feel'; also which characters?"
11	✔	✔		✔		✔	
12	✔	✔					"Remove 'in a group'" "Sporting this and knowing what to do about it are different" "Maybe move to 'Other problems in Com. Skills?"
13	✔	✔	✔	✔	✔	✔	
14	✔	✔	✔	✔		✔	
15	✔						"Change 'masking' to 'hiding'"
16							"This item uses idiomatic language that may be confusing to individuals taking the test." "Eliminate 'white' then might be okay" "This statement could be true of anyone, not just A.D.[AS]"
17	✔	✔	✔	✔		✔	

Table 4.11.2 All Ratings for Stereotyped and Restricted Repetitive Patterns of Interests and Activities

	Method						Comments
Item	*Median*	*Item Ambiguity*	*% Agreement*	*CVI*	*CVR*	VI_K	
1	✔	✔	✔	✔	✔	✔	
2							"Multitasking or being able to shift focus easily?" "Irrelevant"
3	✔	✔					"Consider rephrasing (e.g., "I enjoy focusing on details"
4							"Awkward wordings"
5		✔					"[Could change to] 'I like to wear my favourite clothes almost everyday if I could'" "I like to wear the same clothes everyday"
6	✔	✔					"Consider rewording for clarity and simplicity. Break down into more than one item"
7	✔	✔		✔		✔	"Maybe expand beyond 'list of things'" "Remove 'even when'"
8	✔	✔	✔	✔	✔	✔	
9	✔	✔	✔	✔	✔	✔	
10	✔	✔	✔	✔	✔	✔	"'comfortable' instead of 'unease'"
11		✔					"This item refers to an awareness of how routines would bother other people, which may not be appropriate for an individual with AS who cannot empathize with another individual's feelings."
12	✔	✔	✔	✔	✔	✔	
13	✔	✔	✔	✔	✔		
15	✔	✔		✔		✔	"Consider separating into two items: One that refers to strong interests, and another that refers to becoming upset when those interests cannot be pursued."
16		✔					"Specific fascination won't be applicable to everyone"
17	✔	✔		✔		✔	
18	✔	✔					
19	✔	✔	✔	✔	✔	✔	

Table 4.11.3 All Ratings for Sensorimotor and Stereotyped Patterns of Behaviours

	Method						Comments
Item	*Median*	*Item Ambiguity*	*% Agreement*	*CVI*	*CVR*	VI_K	
1		✔					"Awkward wording" "What do you mean by 'postures'"
2	✔	✔	✔	✔	✔	✔	
4		✔					"Any sort of touching?"
5	✔	✔	✔	✔			"I like how certain foods feel in my mouth"
6	✔	✔	✔	✔	✔	✔	
7	✔	✔		✔			
8	✔	✔	✔	✔		✔	
9	✔	✔	✔	✔			"Some texture bothers me a lot" "Clarify 'textures'"
10	✔						"Consider rewording to remove reference to 'painful noises' and 'high-pitched noise'; the reactions described in these items may not necessarily imply a hyper-sensitivity to noise." "Why just high-pitched"
11							"I don't think prosody goes here" "Unusual is too broad, needs to be better defined. Maybe move to communication"
12	✔	✔				✔	"Consider rewording to remove reference to 'painful noises' and 'high-pitched noise'; the reactions described in these items may not necessarily imply a hyper-sensitivity to noise." "Maybe re-word to reflect over-sensitivity to sounds that others are comfortable with"
13	✔	✔	✔	✔	✔	✔	
14	✔	✔	✔	✔	✔	✔	
15	✔	✔	✔	✔		✔	
16	✔	✔	✔	✔	✔	✔	
17	✔						"I often find myself engaging in inappropriate behaviours"

Table 4.11.4 All Ratings for Deficit in Social Communication Skills: Problems in Pragmatic Language

	Method						Comments
Item	*Median*	*Item Ambiguity*	*% Agreement*	*CVI*	*CVR*	VI_K	
1	✔	✔	✔	✔		✔	"I have trouble staying on topic in a conversation"
2	✔	✔	✔	✔	✔	✔	
3	✔	✔	✔	✔	✔	✔	
4	✔	✔	✔	✔	✔	✔	
5	✔	✔					"Remove negative language"
6	✔	✔	✔	✔		✔	
7	✔	✔	✔	✔	✔	✔	
8	✔	✔	✔	✔	✔	✔	
9	✔	✔	✔	✔	✔	✔	
10	✔	✔	✔	✔	✔	✔	
11	✔	✔	✔	✔	✔	✔	
12	✔	✔	✔	✔	✔		
13	✔	✔	✔	✔	✔	✔	
14	✔	✔				✔	
15	✔	✔	✔	✔		✔	
16	✔	✔	✔				"Remove reference to people'
18	✔	✔		✔		✔	"I am aware when I talk loudly"
19	✔	✔	✔	✔	✔	✔	

Table 4.11.5 All Ratings for Deficit in Social Communication Skills: Other Problems in Communication Skills

	Method						Comments
Item	*Median*	*Item Ambiguity*	*% Agreement*	*CVI*	*CVR*	VI_K	
1	✔						
2	✔	✔	✔	✔		✔	
3	✔	✔	✔	✔			"Consider including additional items to capture loneliness" "I am frustrated because I have no friends"
4	✔	✔					"Perhaps this item would benefit from more clarity"
6	✔	✔	✔	✔	✔	✔	
7		✔					
8	✔	✔	✔	✔	✔	✔	
9	✔	✔	✔	✔		✔	
10	✔	✔	✔	✔	✔	✔	
11	✔	✔	✔	✔			
12	✔	✔	✔	✔		✔	
13	✔	✔	✔	✔	✔	✔	
14	✔	✔	✔	✔	✔	✔	
15	✔	✔	✔	✔	✔	✔	
16	✔	✔	✔	✔	✔	✔	
17	✔	✔	✔	✔		✔	"Maybe better suited for Empathy domain"
18	✔	✔	✔	✔		✔	"Too general"
19	✔	✔	✔	✔	✔	✔	

4.4.2.5 Item difficulty. Some of the items that have been commented on by the expert panel members are items that may be too difficult for individuals with AS to answer. For example, "I can always put myself in others' shoes" and "I would rather tell a white lie than hurt someone's feelings" in the Empathy. People with AS may have difficulty interpreting the underlying meaning of "white lie" and "putting myself in others' shoes". Therefore, these items may need further revision to fit the target population.

4.4.2.6 Suggestions on item deletion. Due to the few items that met all requirements and after examining the comments made by judges, it was determined that items meeting the criteria of less than 4 methods (66.67% agreement) should be removed. Many of the comments with at least 4 methods of agreement were related to choice of wording but not substantive content, and therefore setting the cut-off at this number appears to be reasonable.

Using this new criteria, there are seven items remaining for the domain EMP (Item 1, 4, 8, 11, 13, 14, and 17), ten items for SIA (Item 1, 7, 8, 9, 10, 12, 13, 15, 17, and 19); nine items for SPB (Item 2, 5, 6, 8, 9, 13, 14, 15, and 16); 15 items for DPL (Item 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 15, 18, and 19); and 14 items remaining for DCS (Item 2, 3, 6, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, and 19). These items are listed in Table 4.12 below. Judges comments were included as suggestions for future revision. Table 4.13 categorized the types of comments received from judges.

Table 4.12 Items Remaining after Item Deletion

Empathy		
Item #	**Item**	**Judges' Comments**
1	It is difficult for me to understand others' feelings.	
4	I can easily tell when friends need to be comforted.	
8	I am an understanding person when people tell me their problems.	
11	I am good at predicting how someone will feel.	
13	I have been told that I am good at understanding how others are feeling.	"Perhaps consider including items that only refer to the self, rather than other referents"
14	Other people often say that I am insensitive, though I don't see why.	"Remove 'though I don't see why'"
17	I am considered a compassionate person.	
Stereotyped and Restricted Repetitive Patterns of Interests and Activities		
1	I feel distressed when things do not go as expected.	"This should be in domain but doesn't relate to your definition"
7	I memorize lists of things that interest me, even when they have no practical use.	"Maybe expand beyond 'list of things'" "Remove 'even when'"
8	When I go somewhere, I have to follow a familiar route or I can get very upset.	
9	I like things to be exactly the same day after day.	
10	I have certain routines that I have to follow or I will feel unease.	"'comfortable' instead of 'unease'"
12	Changes to my routine would upset me.	
13	I tend to notice details that others do not.	
15	I tend to have very strong interests, which I get upset about if I can't pursue.	"Consider separating into two items: One that refers to strong interests, and another that refers to becoming upset when those interests cannot be pursued."
17	I notice patterns in things all the time.	
19	I get extremely upset when there is a sudden change of plans.	"Maybe describe behaviour associated with 'extremely upset'"
Sensorimotor and Stereotyped Patterns of Behaviours		
2	I have been told that my behaviours are repetitive.	
5	It is important for me to notice how food feels in my mouth.	"I like how certain foods feel in my mouth"
6	I am very sensitive to the way my clothes feel on my skin.	
8	I am sensitive to smells.	
9	Some textures that do not bother others tend to bother me a lot.	"Some texture bothers me a lot" "Clarify 'textures'"
13	Sometimes things that should feel painful are not (for example, when I hurt myself).	"Remove bracketed information; may require clarity"
14	I often notice small sounds when others do not.	"I am very sensitive to very quiet noises"
15	Spinning around calms me down when I am feeling stressed.	"Add pressure, or other atypical soothing strategies" "Expand beyond just spinning"
16	I can easily get overwhelmed with multiple sensations at the same time.	"I often find myself engaging in inappropriate behaviours"

Deficit in Social Communication Skills: Problems in Pragmatic Language		
1	I have been told that my words are often unrelated to the conversation.	"I have trouble staying on topic in a conversation"
2	I am often told not to interrupt when others are talking.	"What if no one tells them but they do it a lot?"
3	I am often told that I should look at the people I am talking to.	
4	Sometimes I am not aware that my words have made people feel uncomfortable.	"I am aware when my words have made people feel comfortable"
6	It is difficult for me to initiate conversation.	"Initiate a conversation with someone"
7	I am good at making friendly conversation with people I just met.	
8	I find it difficult to know when it is my turn to talk during a conversation.	"I know when it is my turn to talk during a conversation"
9	I can easily tell if someone is interested in what I am saying.	
10	It is difficult for me to understand social cues (e.g., body language).	
11	I have been told that my facial expressions are often inappropriate.	
12	I find it difficult to know when someone is being polite.	
13	I have difficulty knowing how to keep a conversation going.	"I know how to keep a conversation going"
15	I often have difficulty ending a conversation.	
18	Sometimes I am not aware that I am talking too loudly.	"I am aware when I talk loudly"
19	I have been told that my tone of voice does not change when I speak.	
Deficit in Social Communication Skills: Other Problems in Communication Skills		
2	I prefer to be alone than in a group.	
3	I am frustrated about not having friends.	"Consider including additional items to capture loneliness" "I am frustrated because I have no friends"
6	I would not easily change my behaviour even if others were present.	
8	I find conflict management difficult.	
10	I understand the reason for others' need of privacy.	
11	I would rather go out alone than with someone I know.	
12	Meeting new people is usually easy for me.	
13	I feel comfortable being in a social situation with others.	"I feel comfortable being around others"
14	It is almost impossible for me to function in groups.	"I enjoy being in a group of people"
15	I am considered a loner by those who know me best.	"I spend most of my time alone" "No need for 'best'"
16	It is difficult to figure out what other people expect of me.	
17	It is hard for me to see why some things upset people so much.	"Maybe better suited for Empathy domain"
18	New situations make me anxious.	"Too general"
19	I enjoy meeting new people.	

Table 4.13 Categorization of Judges' Comments

Comment Type	Examples of Comments
Clarify	"Remove bracketed information; may require clarity"
Content Related	"Perhaps consider including items that only refer to the self, rather than other referents"
Complexity	"Consider separating into two items: One that refers to strong interests, and another that refers to becoming upset when those interests cannot be pursued."
Change of Wording	"'Comfortable' instead of 'unease'"

4.4.2.7 Comparison of analytical methods. As mentioned above, the item ambiguity seems to be one of the most generous of all measures, in which 13 items in EMP, 16 items in SIA, 13 items in SPB, 18 items in DPL, and 17 items on DCS satisfied the condition of low ambiguity. That is, using this method, more than 81% of the items were considered as "meeting the criteria". The median method also appears to be lenient identifying 14 items in EMP, 13 items in SIA, 13 items in SPB, 18 items in DPL, and 18 items in DCS. The median also agreed with the item ambiguity at over 90% of the time. However, on the other end of the spectrum, the CVR seems to be too strict (identifying only 3 items in EMP, 7 items in SIA, 5 items in SPB, 11 items in DPL, and 8 items in DCS) due to the high .99 acceptance level required to ensure a 95% confidence level for a one-tailed test. If there were more judges (>7) in this study, the acceptance level would have lowered to .75 according to the table provided by Lawshe (1975, p. 567). Therefore, an exploratory study with only a few judges may want to consider using methods other than the CVR, so that more items may be kept for further and better revisions, if possible. Examining the percentage agreement, the CVI and the VI_K, their agreements seem to be quite consistent with one another at the moderate level with about 73% of the time all three methods satisfied the items.

Table 4.14 Characteristic Summary of Analytical Methods

Analytical Methods	Characteristics
Descriptive Analyses	
Item Ambiguity	Too lenient – Method does not discriminate well between items
Median	Too lenient - Method does not discriminate well between items
Percentage Agreement*	Intermediate – Appears to be more balanced
Quantitative Analyses	
CVR	Too conservative – many items were discarded
CVI*	Intermediate – Appears to be more balanced
VI_K*	Intermediate – Appears to be more balanced

* Represents methods that worked best in this study

CHAPTER 5

5. Discussions and Conclusions

This final chapter provides a summary of the purpose of the research, and the procedures utilized in this study. It will be followed by discussions of the results, limitations of the research, and future directions for those interests in expanding on this research. Finally, conclusions of this study will be drawn.

5.1 Summary

5.1.1 Purpose of Research

Based on a thorough literature search, there is currently no existing self-screening instrument for Asperger's Syndrome (AS) for adults. The development of a self-screening instrument would assist individuals who may suspect themselves of having AS, but who have never obtained a diagnosis of AS. A self-screen would also help individuals seeking proper diagnosis if required. To further address the issue that many assessment measures in the field of Autism Spectrum Disorder (ASD) are lacking appropriate psychometric properties, it is important for future measures in the field to obtain solid supporting psychometric evidence. The purpose of this study was to develop and gather content validity evidence for an item pool, which eventually will lead to future development of an AS self-screening scale. Content validity evidence was collected using expert judges specializing in psychometric, ASD, school psychology, and speech-language pathology. This process acted as a determination and confirmation about the quality of items, and whether they belonged to the domains for which they were initially developed. Analyses of content validity were conducted using descriptive and quantitative methods. A second purpose of this study was to compare the descriptive and

quantitative methods of analyzing content validity in order to identify the best judgmental analysis procedure for use in similar studies.

5.1.2 Procedures

Three steps were involved in this research study. The first step involved item writing and categorization of items using existing measures of AS. The second step involved the items being sent out to expert judges. The expert panel members were asked to rate the relevancy of each item to their assigned domains. A total of nine experts participated in the study. The final step involved the analysis of judgment ratings using several different methods of analysis. After analysis of aberrant judges (by examining how well the judges were able to detect the *lie* items in each subscale, and by calculating the level of inter-judge agreement), six experts were retained for further analyses. Descriptive analyses and quantitative analyses were performed to determine the quality of items and how well they fit into the assigned domains. A comparison between the analytical methods was also performed.

5.1.3 Discussion of Findings

Based on a search of the literature and the DSM-IV-TR (APA, 2000) diagnostic criteria of AS, five AS domains were created: Empathy (EMP); Stereotyped and Restricted Repetitive Patterns of Interests and Activities (SIA); Sensorimotor and Stereotyped Patterns of Behaviours (SPB); Deficit in Social Communication Skills: Problems in Pragmatic Language (DPL); and Deficit in Social Communication Skills: Other Problems in Communication Skills (DCS). Following the collection of expert panel member judgments, the median, item ambiguity, percentage agreement, content validity index (CVI), content validity ratio (CVR), and content validity coefficient (VI_K) were calculated in order to analyze the relevancy of the items. The final

results suggested that seven items from the EMP subscale, 10 items from the SIA subscale, nine items from the SPB subscale, 15 items from the DPL subscale, and 14 items from the DCS subscale should be retained. Three of the subscales, the EMP, SIA, and SPB subscales, each contained a number of deleted items, which may suggest either issues with the development of the items such as weak wordings and/or issues with the boundaries of the domain definitions meaning a possible domain underrepresentation when the definitions did not address the domains entirely. As Rogers (2010) suggested that it is extremely important to ensure that the domains have been properly identified and defined before any further procedures begin, better-defined domains may lead to items being judged as more relevant.

A comparison of all methods of judgmental analysis demonstrated that item ambiguity was one of the most generous of all the methods as over 81% of the items satisfied the condition of low ambiguity. The median calculation was at a similar level with the item ambiguity in terms of generosity. It agreed with the item ambiguity at over 90% of the time. In contrast, the CVR was considered extremely conservative due to the high requirement of .99 acceptance level, in which it may benefit greater if there are at least eight judges because the acceptance level will lowered to .75. In this study, fewer items were satisfied under the CVR condition, which suggests that the CVR method may be too conservative a method to use when there are few expert panel members. Researchers may wish to consider using methods other than the CVR when the study is exploratory in nature, and there are less than seven judges, so that more items can be retained for future revisions. At the moderate level, percentage agreement, CVI, and VI_K had more consensuses among each other on the item ratings. All three methods agreed with each other approximately 73% of the time. One may conclude that the percentage agreement is one of the better descriptive analysis methods and the CVI and VI_K two of the better quantitative

analyses, especially when the number of expert panel members is low, or the research is exploratory in nature. In this study, there was no set requirement on the number of retained items required for each domain, as the purpose was to explore the quality of items. However, despite different characteristics of each method of analyzing ratings, *Problems in Pragmatic Language* and *Other Problems in Communication Skills* appear to have many items satisfying all the analytical methods in comparison to other domains. It suggests that these two domains may contain better developed, more relevant, and more representative items.

In terms of selection of judges, this study has demonstrated the importance of ensuring that the judges are experts in the field of the research context. In this case, judges should be chosen as experts in the field of AS and ASD. Examining the background of aberrant judges, two out of the three were Psychometricians, and the third judge was in the field of psychometrics and ASD. It was suspected that the Psychometricians were rating the items using a different perception. They may have been focusing on the item writing rules or the psychometric properties of the items instead of the context of AS. Therefore, it is very important to obtain judges in the field of interest. The inclusion of Psychometricians in this study also demonstrated the importance of having enough judges so that judges can be separated into similar groups for further comparisons on how they rated the items differently.

5.1.4 Limitations of Study

One significant limitation of this study was the low number of expert judge members in the field of ASD. Of the nine judges who chose to participate in the study, only three had associations with the field of ASD. In addition, one of these three experts was determined as an aberrant judge. If more judges were in the related field, then the content validity evidence may be more powerful. A second limitation of this study was the lack of judges from outside of Canada.

Of the nine participants, one participant was from the United States of America. However, the judge from the USA was one of the three aberrant judges, and as such the ratings of this individual were not included in the final analyses. The lack of American judges limits the generalization of this study for use outside of Canada. A third limitation of this study was based on the use of existing items from a combination of current AS measures. Although the measures chosen were the measures with more psychometric evidence, revisions and the editing of contents can be restricted, and some of the items may not represent their assigned domains as well as if they were created specifically for this scale. Sometimes it may be easier to write original items rather than trying to revise poorly written items.

The last limitation was the possible lack of representation of the construct in this item pool and domains. Domains were created based on observable characteristics in AS individuals from the DSM-IV-TR (APA, 2000) diagnostic criteria, however, there could be other *symptoms* of these individuals that were not expressed in the domains. For example, language and cognitive development in these individuals as children cannot be measured in the context of a self-screening instrument, as the majority of adults using the self-report instrument are unlikely to recall their childhood medical history. Therefore, further medical history may be required to correctly identify AS in adults. However, it is important to note that the items in this study were built for a self-screening instrument, in which its intention was not to diagnose people, but rather to assist people to further understand their behaviours and to seek professional help if required.

5.1.5 Future Directions

Further revision of items can be conducted with the items for better quality and representation of domains. In addition, the revision of the EMP, SIA, and SPB domains may need to be conducted due to the smaller number of items representing these domains. More items

can be written for these domains and redefining and clarifying the domain definitions may help to develop better fitting items. The DSM-IV-TR (APA, 2000) criteria for AS characteristics can also be explored for item revision, especially criterion (c) The disturbance causes clinically significant impairment in social, occupational, or other important areas of functioning (APA, 2000, p. 84). It has been noticed that some items in this study may or may not belong to this category, however it cannot be easily determined which qualities would considered *clinically significant* without further explorations and clinical measures. Future research should obtain more judges in each of the fields (at least five experts in each field as proposed in this research) so groups can be formed based on backgrounds of the judges (i.e. content [ASD and related] versus Psychometric judges), and comparisons can be made based on characteristics of these judges. Item quality can also be determined across judges in groups to create higher generalizability of items. Instead of inviting individuals from universities, recruitment of expert panel members can be expanded to clinics, service centers, and agencies specializing in helping AS individuals. Going outside of the university setting would increase the potential number of expert panel members. To further address the problem of generalization of items to populations outside Canada, future research should acquire judges from across North America, so that the selected items rated from these judges can be universally used. In terms of content validity analyses, future research may compare other methods that were not utilized in this study to further "rank" these analysis methods.

Results of this research can be used for further development of the AS self-screening instrument following the remaining steps (steps five to eight) suggested by DeVellis (2003): (5) considering inclusion of validation items; (6) administering items to a development sample; (7) evaluate the items; and (8) optimize scale length. The current items can be combined into one

single self-screening instrument once a scoring system (i.e., response scale) has been established. One should note that this research only provides one type of validity evidence (content validity) for this potential measure. As Messick (1991) suggested that validity evidence should be considered as unitary, and multiple evidences should be collected for a test, additional research is required. Other validity evidence such as construct validity and criterion-related validity evidence should also be collected. In addition to validity evidence, reliability evidence such as test-retest reliability and internal consistency of the items should also be obtained. Continuation of collecting validity evidence can be conducted along with examining reliability issues while performing pilot studies.

5.2 Conclusion

This research developed a potential useful item pool for an AS self-screening instrument with good evidence of content validity. This evidence was determined by using different analyses, in which items receiving the most consensus were retained. A comparison of methods has also shown that in an exploratory study with a few judges, the percentage agreement appears to be a good method to use for descriptive analysis, and the CVI and VI_K methods appear to be good methods for quantitative analysis. As suggested in future research, more validity evidence in addition to reliability evidence should be collected in order for the instrument to be a truly useful test with proper psychometric properties.

This study provides the resources for better development of assessment measures in the field of AS, in the hope to bring more attention to the importance of appropriate psychometric properties in AS and ASD tests. For benefits in the psychometric field, this research examined the popularly used content validity analyses to shed a light on which methods should be used.

REFERENCES

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement, 45,* 131-142.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* California: Brooks/Cole Publishing Company.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders (4th ed.).* Washington, DC: Author.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders (4th ed.-text revised).* Washington, DC: Author.

American Speech-Language Association (2009). Social Language Use (Pragmatics). Retrieved December 10, 2009, from American Speech-Language Association Online via http://www.asha.org/public/speech/development/pragmatics.htm.

Autism Society Canada. (2005). What is autism spectrum disorder? Terms and definitions. Retrieved October 10, 2008, from Autism Society Canada Online via http://www.autismsocietycanada.ca/understanding_autism/what_are_asds/index_e.html.

Allik, H., Larsson, J., & Smedje, H. (2006). Sleep patterns of school-age children with 98sperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders, 36,* 585-595.

Barnhill, G. P. (2007). Outcomes in adults with asperger syndrome. *Focus on Autism and other Developmental Disabilities, 22 (2),* 116-126.

Baron-Cohen, S. (2001). Theory of mind in normal development and autism. *Prisme, 34,* 174-183.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition, 21,* 37-46.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with 99sperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34 (2),* 163-175.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb, I. (2001). The 'Reading the mind in the eyes' test revised version: a study with normal adults, and adults wit Asperger syndrome or high-functioning autism. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 42,* 241–251.

Baron-Cohen, S., Wheelwright, S., Robinson, J., & Woodbury-Smith, M. (2005). The adults 99sperger assessment (AAA): A diagnostic method. *Journal of Autism and Developmental Disorders, 35,* 807-819.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from 99sperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 11 (1),* 5-17.

Campbell, J. M. (2005). Diagnostic assessment of 99sperger's disorder: A review of five third-party rating scales. *Journal of Autism and Developmental Disorders, 35 (1),* 25-35.

Carrington, S., & Templeton, E. (2003). Adolescents with 99sperger syndrome and perceptions of friendship. *Focus on Autism and other Developmental Disabilities, 18 (4),* 211-218.

Cronbach, L. J., & Shavelson, R. J. (Ed.). (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64(3),* 391-418.

Davis, M. (1980). A multidimensional approach to individual differences in empathy. *Catalog of*

Selected Documents in Psychology, 10, 85.

DeVellis, R. F. (2003). *Scale development: Theory and applications.* United Kingdom: Sage Publications Ltd.

Dorris, L., Espie, C.A.E., Knott, F., & Salt, J. (2004). Mind-reading difficulties in the siblings of people with 100sperger's syndrome: evidence for a genetic influence in the abnormal development of a specific cognitive domain. *Journal of Child Psychology and Psychiatry, 45,* 412-418.

Ehlers, S., & Gillberg, C. (1993). The epidemiology of 100sperger syndrome. A total population study. *Journal of Child Psychology and Psychiatry, 34 (8),* 1327-1350.

Frey, B. B., Petersen, S., Edwards, L. M., Pedrotti, J. T., & Peyton, V. (2005). Item-writing rules: Collective wisdom. *Teaching and Teacher Education, 21,* 357-364.

Frisbie, D. A. (1988). Reliability of scores from teacher-made tests. *Instructional Topics in Educational Measurement, module 3,* 55-65.

Frombonne, E., & Tidmarsh, L. (2003). The epidemiology of 100sperger disorder. *Child and Adolescence Clinics of North America, 12 (1),* 15-21.

Gallucci, G., Hackerman, F., & Schmidt, C. W. (2005). Gender identity in an adult male with 100sperger's syndrome. *Sexuality and Disability, 23 (1),* 35-40.

Ghaziuddin, M. (2005). A family history study of 100sperger syndrome. *Journal of Autism and Developmental Disorders, 35 (2),* 177-182.

Gillberg, C., & Billstedt, E. (2000). Autism and 100sperger syndrome: coexistence with other clinical disorders. *Acta Psychiatrica Scandinavica, 102,* 321-330.

Gillberg, C., Gillberg, C., Rastam, M., & Wenta, E. (2001). The 101sperger syndrome (and high-functioning autism) diagnostic interview (ASDI): a preliminary study of a new structured clinical interview. *Autism, 5(1),* 57-66.

Gilliam, J. E. (2001). *Gilliam Asperger Disorder Scale (GADS).* Austin, TX: Pro-ed.

Harvill, L. M. (1991). Standard error of measurement. *Instructional Topics in Educational Measurement module 9,* 181-189.

Haynes, S. N., Richard, C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7 (3),* 238-247.

Hellsten, L. M. (2008). Accumulating content validity evidence: Assessing expert panel ratings of item relevance and representativeness. Presented at the 2008 National Council on Measurement in Education Annual Conference, New York, NY, March 25, 2008.

Hellsten, L. M. (2009a). Educational Psychology 844: Accumulating Evidence of Content Validity. Unpublished Presentation, University of Saskatchewan.

Hellsten, L. M. (2009b) Education Psychology 844: Validity and Reliability. Unpublished Presentation, University of Saskatchewan.

Hendry, C. N. (2000). Childhood disintegrative disorder: Should it be considered a distinct diagnosos? *Clinical Psychology Review, 20 (1),* 77-90.

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21 (5),* 967-988.

Hoelter, J.W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods and Research, 11*. 325-344.

Holm, V. A. (1985). Rett syndrome: A case report from an audiovisual program. *Brain and*

Development, 7, 297-299.

Howlin, P. (2000). Assessment instruments for 102sperger syndrome. *Child Psychology & Psychiatry Review, 5 (3),* 120-127.

Hurlbutt, K., & Chalmers, L. (2004). Employment and adults with 102sperger syndrome. *Focus on Autism and other Developmental Disabilities, 19 (4),* 215-222.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69,* 85-98.

Klin A., Volkmar F.R., , Sparrow S.S., , Cicchetti D. V., and Rourke B.P. (1995) Validity and neuropsychological characterization of Asperger Syndrome: Convergence with Nonverbal Learning Disabilities Syndrome. Journal of Child psychology and Psychiatry 36, 1127-40.

Koegel, A. K. (2008). Evidence suggesting the existence of 102sperger syndrome in the mid-1800s. *Journal of Positive Behavior Intervention, 10 (4),* 270-272.

Kohler, W. (1929). *Gestalt psychology.* New York: Liveright.

Kraemer, B., Delsignore, A., Gundelfinger, R., Schnyder, U., & Hepp, U. (2005). Comorbidity of 102sperger syndrome and gender identity disorder. *European Child & Adolescent Psychiatry, 14,* 292-296.

Kuusikko, S., Pollock-Wurman, R., Jussila, K., Carter, A. S., Mattila, M., Ebeling, H., Pauls, D. L., & Moilanen, I. (2008). Social anxiety in high-functioning children and adolescents with autism and 102sperger syndrome. *Journal of Autism and Developmental Disorders, 38,* 1697-1709.

Landa, R., Piven, J., Wzorek, M. M., Gayle, J. O., Chase, G. A., & Folstein, S. E. (1992). Social language use in parents of autistic individuals. *Psychological Medicine, 22,* 245-254.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: reliability and validity of the empathy quotient. *Psychological Medicine, 34*, 911-924.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28,* 563-575.

Lawson, J., Baron-Cohen, S., & Wheelwright, S. (2004). Empathising and systemizing in adults with 103sperger syndrome. *Journal of Autism and Developmental Disorders, 34,* 301-310.

Lim, H. K., & Slaughter, V. (2008). Brief report: Human figure drawings by children with 103sperger's syndrome. *Journal of Autism and Developmental Disorders, 38,* 988-994.

Liu, X., Hubbard, J. A., Fabes, R. A., & Adam, J. B. (2006). Sleep disturbance and correlates of children with autism spectrum disorders. *Child Psychiatry and Human Development, 37,* 179-191.

Lopata, C., Thomeer, M. L., Volker, M. A., & Nida, R. E. (2006). Effectiveness of a cognitive-behavioral treatment on the social behaviors of children with 103sperger disorder. *Focus on Autism and other Developmental Disabilities, 21 (4),* 237-244.

Lopez, B. R., Lincoln, A. J., Ozonoff, S., & Lai, Z. (2005). Examining the relationship between executive functions and restricted, repetitive symptoms of autistic disorder. *Journal of Autism and Developmental Disorders, 35 (4),* 445-460.

Lu, K. H. (1971). A measure of agreement among subjective judgments. *Educational and Psychological Measurement, 31,* 75-84.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35, (6),* 382-385.

Manjiviona, J., & Prior, M. (1995). Comparison of 104sperger syndrome and high-functioning autistic children on a test of motor impairment. *Journal of Autism and Developmental Disorders, 25 (1),* 23-39.

Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and Language, 85,* 451-466.

Mayes, L., Volkmar, F., Hooks, M., & Cicchetti, D. (1993). Differentiating pervasive developmental disorder not otherwise specified from autism and language disorders. *Journal of Autism and Developmental Disorders, 23 (1),* 79-90.

McDermott, P. A., & Watkins, M. W. (1979). A program to evaluate general and conditional agreement among categorical assignments of many raters. *Behaviour Research Methods and Instrumentation, 11,* 399-400.

Messick, S. (1990). Validity of test interpretation and use. Princeton, NJ: Educational Testing Service.

Meyer, R. N. (2000). Adult version: Australian scale for 104sperger syndrome. Retrieved March 23, 2009, from ASPIRES Online via http://www.aspires-relationships.com/articles_adult_version_austrailian_scale_for_as.htm.

Miller, J. M. & Penfield, R. D. (2005). Using the score method to construct asymmetric confidence intervals: An SAS program for content validation in scale development. *Behavior Research Methods, 37,* 450-452.

Murphy, D. (2006). Theory of mind in 104sperger's syndrome, schizophrenia and personality disordered forensic patients. *Cognitive Neuropsychiatry, 11(2),* 99-111.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ozonoff, S., Rogers, S.J., & Hendren, R.L. (Eds.). (2003). *Autism spectrum disorders*: A research review for practitioners. Arlington: American Psychiatric Publishing.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendation. *Research in Nursing & Health, 30,* 459-467.

Popham, W. J. (1978). *Criterion-referenced measurement.* Englewood Cliffs, NJ: Prentice-Hall.

Quitkin F, Rifkin A, Klein DF. (1976). Neurologic soft signs in schizophrenia and character disorders. *Arch. Gen. Psychiatry (33),* 845–853.

Rinehart, N. J., Bradshaw, J. L., Brereton, A. V., & Tonge, B. J. (2002). Lateralization in individuals with high-functioning autism and 105sperger's disorder: A frontostriatal model. *Journal of Autism and Developmental Disorders, 32 (4),* 321-332.

Ritvo, R. A., Ritvo, E. R., Guthrie, D., Yuwiler, A., Ritvo, M. J., & Weisbender, L. (2008). A scale to assist the diagnosis of autism and 105sperger's disorder in adults (RAADS): A pilot study. *Journal of Autism and Developmental Disorders, 38,* 213-223.

Robinson, R. (1998). Obsessive compulsive disorder in children and adolescents. *The British Journal of Psychiatry, 35,* 91-96.

Rogers, K., & Dziobek, I. (2007). Who care? Revising empathy in 105sperger syndrome. *Journal of Autism and Developmental Disorder, 37,* 709-715.

Rogers, W. T. (2010). Educational Psychology 507: The Nature of Validity. Unpublished Manuscript, University of Alberta.

Rovinelli, R. J. & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal For Educational Research, 2, 2,* 49-60.

Schatz, A. M., Weimer, A. K., & Trauner, D. A. (2002). Brief report: Attention differences in 106sperger syndrome. *Journal of Autism and Developmental Disorders, 32 (4),* 333-336.

Shamay-Tsoory, S. G. (2008). Recognition of 'fortune of others' emotions in 106sperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorder, 38,* 1451-1461.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10,* 640-656.

Stoesz, B. M., Montgomery, J. M., & Hellsten, L. M. (2010). *Evaluation of five instruments for the assessment of asperger's disorder in adults.* Manuscript submitted for publication.

Tani, P., Linderg, N., Appelberg, B., Wendt, T. N., Wendt, L., & Anen, T. P. (2006). Clinical neurological abnormalities in young adults with 106sperger syndrome. *Psychiatry and Clinical Neuroscience, 60,* 253-255.

Tinsley, H.E., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology, 22,* 358-376.

Titchener, E. (1909). *Elementary psychology of the thought processes.* New York: Macmillian.

Thede, L. L., & Coolidge, F. L. (2007). Psychological and neurobehavioral comparisons of children with 106sperger's disorder versus high-functioning autism. *Journal of Autism and Developmental Disorder, 37,* 847-854.

Traub, R. E., & Rowley, G. L. (1991). Understanding reliability. *Instructional Topics in Educational Measurement, module 8,* 171–197.

Trevanthan, E., & Naidu, S. (1988). The clinical recognition and differential diagnosis of Rett syndrome. *Journal of Child Neurology,* 3(Suppl.), S6-S16.

Tryon, P. A., Mayes, S. D., Rhodes, R. L., & Waldo, M. (2006). Can 107sperger's disorder be differentiated from autism using DSM-IV criteria? *Focus on Autism and Other Developmental Disabilities, 21 (1),* 2-6.

Van Acker, R. (1991). Rett Syndrome: A review of current knowledge. *Journal of Autism and Developmental Disorders, 21 (4),* 381- 406.

Waltz, C. W. & Bausell, R. B. (1983). *Nursing research: Design, statistics, and computer analysis.* Philadelphia, PA: F. A. Davis Company.

Wing, L. (1981). Asperger syndrome: A clinical account. *Psychological Medicine, 11*, 115-129.

Wing, L., & Gould J. (2006). Letter to the editor: Comment on "the adult 107sperger assessment (AAA): A diagnostic method", simon baron-cohen, sally wheelwright, 107sperg robinson, and marc woodbury-smith, journal of autism and developmental disorders vol 35(6). *Journal of Autism and Developmental disorders, 36,* 1141-1142.

Witt-Engerstrom, I. (1987). Rett syndrome: A retrospective pilot study on potential early predictive symptomatology. *Brain and Development, 9,* 481-486.

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for 107sperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders, 35 (3),* 331-335.

World Health Organization. (1993). *International statistical classification of diseases and related health problems (10th revision).*

Wynd, C. A., Schmidt, B., & Schaefer, M.A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research, 25,* 508-518.

Zandt, F., Prior, M., & Kyrios, M. (2007). Repetitive behavior in children with high function autism and obsessive compulsive disorder. *Journal of Autism and Developmental Disorder, 37*, 251-259.

APPENDIXES

Appendix A Diagnostic Instruments

Table A1

Examples of Existing ASD Measures

ASD Measures	**Authors**
• Autism Diagnostic Observation Schedule – Generic (ADOS-G)	• Lord, Risi, Lambrecht, Cook Jr., Leventhal, DiLavore, Pickles & Rutter (2000)
• Childhood Autism Rating Scale (CARS)	• Schopler, Reichler, & Renner (1988)
• Gilliam Autism Rating Scale (GARS)	• Gilliam (1995)
• Autism Diagnostic Interview – Revised (ADI-R)	• Lord, Rutter, & LeCouteur (1994)
• Autism Behaviour Checklist (ABC)	• Krug, Arick, & Almond (1980)

Table A2

Examples and Descriptions of AS/HFA Measures

AS/HFA Measures	**Authors**	**Age Group**	**Purchase Required**
The High-Functioning Autism Spectrum Screening Questionnaire (ASSQ)	Ehlers, Gillberg, & Wing (1999)	Children and Adolescents	
*Ritvo Autism and Asperger's Diagnostic Scale (RAADS)	Ritvo, Ritvo, Guthrie, Yuwiler, Ritvo, & Weisbender (2008)	Adults	
*Autism-Spectrum Quotient (AQ)	Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley (2001)	Adults (adults version)	
Asperger Syndrome Diagnostic Scale (ASDS)	Myles, Bock, & Simpson (2001)	5-18 years	✓
*Adult Asperger Assessment (AAA)	Baron-Cohen, Wheelwright, Robinson, & Woodbury-Smith (2005)	Adults	
*The Asperger Syndrome (and High Functioning Autism) Diagnostic Interview (ASDI)	Gillberg, Gillberg, Rastam, &Wentz (2001)	Adolescents and Young Adults	

*Gilliam Asperger's Diagnostic Scale (GADS)	Gilliam (2001)	3-22 years	✓
Krug Asperger's Disorder Index (KADI)	Krug & Arick (2003)	Children	✓
*Australian Scale for Asperger Syndrome – Adult version	Meyer (2000) – adult version; from The Australian Scale for Asperger Syndrome, Attwood (1998) – modified version from Garnett & Attwood (1995)	Adult	
The Australian Scale for Asperger Syndrome	Garnett, & Attwood (1995)	Children	

* = Scales reviewed in Chapter 2

Table A3

Diagnostic Tools Commonly Used

Diagnostic Tools Commonly Used	**Authors**
DSM-IV	APA; American Psychiatric Association (1994)
ICD-10	WHO; World Health Organization (1992)

Appendix B 40 Item-Writing Suggestions by Frey et al. (2005)

1. ‘‘All of the Above’’ should not be an answer

2. ‘‘None of the Above’’ should not be an answer option

3. All answer options should be plausible

4. Order of answer options should be logical or vary

5. Items should cover important concepts and objectives

6. Negative wording should not be used

7. Answer options should include only one correct answer

8. Answer options should all be grammatically consistent with stem

9. Specific determiners (e.g., always, never) should not be used

10. Answer options should be homogenous

11. Stems must be unambiguous and clearly state the problem

12. Correct answer options should not be the longest answer option

13. Answer options should not be longer than the stem

14. Items should use appropriate vocabulary

15. In fill-in-the-blank items, a single blank should be used, at the end

16. Items should be independent of each other

17. In matching, there should be more answer options than stems

18. All parts of an item or exercise should appear on the same page

19. True-false items should have simple structure

20. True-false items should be entirely true or entirely false

21. There should be 3–5 answer options

22. Answer options should not have repetitive wording

23. Point value of items should be presented

24. Stems and examples should not be directly from textbook

25. Matching item directions should include basis for match

26. Answer options should be logically independent of one another

27. Directions should be included

28. Questions using the same format should be together

29. Vague frequency terms (e.g., often, usually) should not be used

30. Multiple-choice stems should be complete sentences

31. There should be an equal number of true and false statements

32. True-false statements should be of equal length

33. Individual items should be short

34. Answer options should be available more than once

35. Number of answer options should be < 7 for elementary age tests

36. Number of answer options should be <17 for secondary age tests

37. Complex item formats (‘‘a and b, but not c’’) should not be used

38. All items should be numbered

39. Test copies should be clear, readable and not hand-written

40. Stems should be on the left, and answer options on the right

Appendix C Methods of Content Validity

General Methods of Analyzing Content Validity Evidence	
Qualitative Methods Descriptive Methods Quantitative Methods - Judgmental Analysis - Generalizability Theory - Factor Analysis - Structural Equation Modeling - Item Response Theory	
Specific Methods of Assessing Judgmental Analysis	
Methods of Analyzing Judgmental Results	
Qualitative Content Analysis Interview Focus group Open-ended feedback	
Descriptive Methods	
Simple Statement of Agreement	
Inter-rater Agreement	
Item Ambiguity *	Rogers (2010)
Mean Item Ratings	
Median Item Ratings *	Rogers (2010)
Percent Agreement *	McDermitt & Watkins, 1979
Specific Quantitative Methods	
Average Congruency Percentage	Popham, 1978
Asymmetric Confidence Interval for the Mean	Miller & Penfield, 2005
Coefficient of Agreement	Lu, 1971
Content Validity Coefficient ($VI_{K)}$ *	Aiken, 1985
Content Validity Index (CVI) *	Waltz & Bausell, 1981; Lynn, 1986
Content Validity Ratio (CVR) *	Lawshe, 1975
Factorial Validity Index (FVI)	Rovinelli & Hambleton, 1977
Index of Item Congruence	Wynd et al., 2003
Latent Partition Analysis	Tinsley & Weiss, 1975
Multi-rater Kappa Coefficient	James et al., 1984

Note. From "Accumulating Content Validity Evidence: Assessing Expert Panel Ratings of Item Relevance and Representativeness," by L. M. Hellsten, 2008, presented at the 2008 National Council on Measurement in Education Annual Conference, New York, NY, March 25, 2008.

* = Methods will be used in current research

Appendix D Modification of Items: Original Items and their Associated Measures

Items removed due to repeat

Remove due to unfit/inappropriateness of language level or content.

Unfit into domains

Measure:
Australian Scale for Asperger's Syndrome – Adult Version (Meyer, 2000; Original children version by Attwood, 1996)

Original Version	***Original Domain***	***Modified: Double-Barrels Removed***	***Modified: First Person***	***Other Modifications on Item Quality***
1. Does the person lack understanding of how to play adult games with others?	Social and Emotional Abilities	-	I have difficulty on understanding how to play adults games wither others.	Understanding game rules is always difficult for me.
2. During unstructured time such as work breaks and informal social events, does he avoid social contact? For example, eats alone, reads, or continues to work.	Social and Emotional Abilities	During unstructured time such as work break, does he avoid social contact?	During unstructured time such as work break, I would avoid social contact.	I enjoy being alone during the break time at work.
3. Is the person unaware of social conventions or codes of conduct including unwritten rules at work? Does he make inappropriate comments of actions? For example, is he unaware of the offending or other unintended effect of his comments?	Social and Emotional Abilities	Is the person unaware of social convention rules at work? Does he make inappropriate comments of actions?	I find it difficult to understand social convention rules. I tend to make inappropriate comments of actions	I find it difficult to understand social norms. I tend to make comments that people find inappropriate.
4. Does the person lack empathy, i.e., an intuitive understanding of another person's feelings? For example, is he not likely to offer an apology or acknowledge his responsibility for a relationship that has failed?	Social and Emotional Abilities	Does the person lack empathy?	It is difficult for me to understand others' feelings.	-
5. Does the person expect other people to know his thoughts, experiences and opinions? For example, he doesn't realize that you couldn't know about something because you were not there at the time. Does he presume you know what you are thinking when you are in the same physical place but your attention has not been directed to him?	Social and Emotional Abilities	Does the person expect other people to know his thoughts? Does the person expect other people to know his experiences? Does the person expect other people to know his opinions?	I expect other people to know my thoughts. I expect other people to know my experiences. I expect other people to know my opinions.	- I expect other people to know what I have experienced. I expect other people to know my opinions on things.
6. Does the person worry excessively or fret about things that change or that don't go as expected? Does the person demand frequent reassurance that matters are OK?	Social and Emotional Abilities	Does the person worry excessively? Does the person fret about things that do not go as expected? Does the person demand frequent reassurance that matters are OK?	I worry about things a lot. I fret about things that do not go as expected. I seek frequent reassurance that things will okay.	- I feel distress when things do not go as expected. -
7. Does the person express concern about loneliness or a tendency to self-isolate? Is he frustrated or anxious about not having any friends or only a few friends? Does he say that he does not know how to make friends?	Social and Emotional Abilities	Does the person express concern about loneliness? Does the person express a tendency to self-isolate? Does the person feel frustrated about not having friends? Does the person say he/she does not know how to make friends?	I am concern about my loneliness. I prefer to be alone than being in a group. I feel frustrated about not having friends. It is difficult for me to make friends.	- - I am frustrated about not having friends. Making friends is difficult for me.

8. Does the person express emotions bluntly? Does he "blurt out" his emotional expressions in ways out of scale to the situation, or before or after emotional expression is expected?	Social and Emotional Abilities			
9. Does the person mean to express one emotion but actually express another?	Social and Emotional Abilities	-	It is difficult for me to express appropriate emotions to others.	It is difficult for me to know what emotions to display to others.
10. Does the person have an unusual attitude towards competition? For example, is he aversive to competitive activities such as sports, games or workplace performance contests? Does he act competitively in activities that call for collaboration and cooperation?	Social and Emotional Abilities	Does the person have an unusual attitude towards competition? Is the person aversive to competitive activities? Does the person act competitively in activities that call for cooperation?	 I do not enjoy activities that require competitions. I tend to play competitively in activities that call for cooperation.	 I enjoy competitive activities. I tend to play competitively in all kinds of activities.
11. Does the person demonstrate indifference to normal or expected peer pressure? Is he generally unaware of widespread crazes or fashions of the moment?	Social and Emotional Abilities	Does the person demonstrate indifference to peer pressure? Is the person generally unaware of widespread fashions of the moment?	 I am usually unaware of the fashion trends.	 -
12. Does the person vocalize exceptionally strong approval or disapproval of benign acts and the choice behaviors of others? Does he attempt to impose his choice as "the only choice" in situation allowing optional choices by others? For example, food preferences, cultural activities, techniques of task performance, and social entertainment.	Social and Emotional Abilities	-	I always try to tell others that my choice is the only choice that they should go with.	-
13. Is the person unaware of others' different styles of learning? If he is aware, is he intolerant of styles other than his own?	Social and Emotional Abilities	Is the person unaware of others' different styles of learning? Is the person intolerant of learning styles other than his own?	 I cannot tolerate others' different style of learning.	 -
14. Does the person not easily modify his behavior while in the presence of persons of different rank or strangers? For example, choice of attire, posture and gestures, addressing the other(s)by their first name(s), making inappropriate demands on the time of others, and not being aware of the differing deference conduct of others. If he is aware, is he critical of that conduct? Does he expect everyone to accept him just as he is?	Social and Emotional Abilities	Does the person not easily modify his behavior while in the presences of persons of rank or strangers?	I would not easily change my behavior even others are present.	-
15. Does the person have difficulty accepting criticism, correction, and direction? Does he have a problem offering the same to others?	Social and Emotional Abilities	Does the person have difficulty accepting criticism? Does the person have difficulty accepting corrections?	It is difficult for me to accept criticisms.	-
16. Does the person have difficulty managing conflict, disagreement, and negotiation? Does the person have trouble with social problem-solving	Social and Emotional Abilities	Does the person have difficulty managing conflict?	I find conflict management difficult.	-

behavior? Does he confront others over differences, pout, or withdraw from the situation in an untimely way rather than remain in uncomfortable or difficult situations?		Does the person have trouble with social problem-solving behavior?		
17. Does the person understand the reason for physical boundaries, personal space, and others' needs for privacy?	Social and Emotional Abilities	Does the person understand the reason for personal space? Does the person understand the reason for others' needs for privacy?	I understand why people need personal space. I understand the reason for others' needs of privacy.	- -
18. Does the person report life-long issues with explosive anger, rage, and lingering resentment over ancient slights?	Social and Emotional Abilities	-	It is difficult for me to control my anger.	-
19. Does the person seen to function in ways suggesting a constant low level of depression?	Social and Emotional Abilities			
20. Does the person take words, phrases, or directions literally? Does he not understand figures of speech and common clichés without explanation? Does he not understand sarcasm? Does he have trouble understanding humor? Does he have an unusual sense of humor? Does he not understand the function of banter and small talk?	Communication Skills	Does he not understand figures of speech? Does he not understand sarcasm? Does he have trouble understanding humor?	Figures of speech are difficult to understand for me. I do not understand sarcasm. I have trouble understanding humor.	- I can easily tell when someone is being sarcastic. It is difficult for me to understand humor.
21. Does the person have an unusual tone of voice (monotone, sing-song an "affected" foreign accent, unusual inflections, prosody, and other oddities of fluid speech? Does he speak with an unusually loud or soft volume level? Does he use changes of tone, inflection, or volume levels appropriate for different levels of conversational formality, location, and topic choice?	Communication Skills		I have been told that my speech is monotone. I have been told that I have an "affected" foreign accent.	
22. When talking, does the person appear uninterested in your side of the conversation? Does he speak in a monologue, exert inappropriate control over the flow or subject matter of conversation, otherwise not consistently engage in a fluid, reciprocal exchange? Is he uncomfortable with pauses or silence in paired or group conversation? Does he fail to observe turn taking rules? Does he interrupt others despite repeated correction? Does he have trouble closing a conversation? Does he miss gestural, postural and facial cues of boredom, agreement, dissatisfaction, impatience and intention to end conversation?	Communication Skills	Does he speak in a monologue? Is he uncomfortable with silence in a group conversation? Does he interrupt others despite repeated correction. Does he have trouble closing a conversation?	Sometimes I find myself speaking in a monologue. I feel uncomfortable with silence in a conversation with others. I am often told not to interrupt when others are talking. I usually have difficulty closing a conversation.	- - -
23. Is the person prone to "go off on tangents" and otherwise become distracted by a minor topic? For example, if there is an agreed-upon agenda, does he have trouble sticking to it? Does he return to something already discussed "for one last word"?	Communication Skills			

24. While directly engaged in conversation, does the person use less eye contact than you would expect? Does he appear to either stare away from you or have a vacant expression when listening or talking? Do his facial gestures, body posture and stance project messages different than his words?	Communication Skills	While directly engaged in conversation, does the person use less eye contact than you would expect?	I have been told many times that I should maintain eye contacts with others during a conversation.	-
25. Is the person's speech over-precise, pedantic, or "professor-ish"? Does he tend to challenge or correct the word choices of others?	Communication Skills	Is the person's speech very "professor-ish"? Does he tend to correct the word choices of others?	I have been told that my speech is very "professor-ish". I tend to correct the word choices of others.	- -
26. Once started, does the person demonstrate an encyclopedic knowledge of a topic? Do you have the sense that regardless of your interest, he starts conversations just to talk about his own? Regardless of what is being discussed, does he repeatedly return to his topic of interest?	Communication Skills			
27. Does the person have difficulty summarizing or "getting to the gist" when reporting conversations or describing events? Does he "ramble" without focus?	Communication Skills	Does the person have difficulty summarizing a conversation?	It is difficult for me to summarize a conversation.	-
28. Does the person say that others characterize him negatively as "a know-it-all"?	Communication Skills	-	Other people called me "a know-it all".	-
29. Does the person have problems repairing a conversation? For example, when he is confused or has lost the train of thought, has he earlier failed to check in to track whether he is "on the same page"? Once confused, does he ask for clarification or redirection? Does he abruptly switch to a different topic without using transition phrases?	Communication Skills		I would ask for clarifications if I were confused during a conversation.	-
30. Does the person engage in audible self-talk during a conversation? When asked questions, does he offer responses that seem repetitive, scripted or askew given the context and the topic of conversation? Does he take an unusually long time responding in conversational give and take?	Communication Skills	Does he take an unusually long time responding in conversational give and take?	It usually takes me some time to respond in a conversation.	
31. Does the person have specific learning disabilities? Examples are problems with math functions and/or written math problems, a history of precocious reading combined with comprehension issues, slow reading speed, dyslexia, bad handwriting, speech delay and pragmatic language problems, short-term working memory deficits leading to frequent checking and reconfirmation behavior, little understanding of the functions of a given behavior, difficulty in perceiving differences between experiences and adjusting responses in accord with new information, repetitive and dysfunctional study habits and learning behaviors with obvious difficulty stopping or changing them.	Cognitive and Executive Function Skills		I have a history of reading problems. I have a history of comprehension problems. I have a history of dyslexia. I have a history of speech delay. I have a history of pragmatic language problems. I have been told my handwriting is bad.	

32. Does the person report having trouble with understanding and following directions? Does he do things "his way" despite directions and instructions requiring a different approach to task completion or performance?	Cognitive and Executive Function Skills	Does he do things "his way" despite instructions requiring a different approach to task completion?	I like to do things my way even instructions stated otherwise are provided.	-
33. Does the person have problems multi-tasking? Must he complete a single activity before" catching up" to perform others? Does he prefer step-by-step instruction? Does he become agitated when given multiple tasks or directed to change his priorities? Can he describe his style of learning? Does he report frustration or stress when being instructed to learn in ways that do not comport with his learning style?	Cognitive and Executive Function Skills	Does the person have problems with multi-tasking? Does he prefer step-by-step instructions? Can he describe his style of learning? Does he report frustration when being instructed to learn in ways that do not comport with his learning style?	I have trouble with multi-tasking. I always prefer step-by-step instructions. I can describe my style of learning. I feel frustrated when I am instructed to learn in ways that differ than my own style of learning.	I have difficulty multi-tasking. I always prefer step-by-step instructions over multiple steps presented at once. I can easily describe my style of learning. -
34. Does he report major study skills impediments in K-12, post-secondary education or vocational and on-the-job training and instruction?	Cognitive and Executive Function Skills			
35. Does the person become disoriented when presented first with the "big picture" of a job? At that moment, does he not want to understand the function of a given task in a greater scheme?	Cognitive and Executive Function Skills			
36. From reports of others or self-reports, would the person consider himself a "dreamer" or "off in my own world" some or much of the time?	Cognitive and Executive Function Skills	-	I consider myself as a dreamer.	-
37. Does the person have projects or interests that others do not understand regardless of his efforts to explain them?	Cognitive and Executive Function Skills	-	Others find it difficult to understand my interests even I tried very hard to explain.	-
38. Does the person have limited interests? Would the person characterize most of his interests as "technical" rather than "artistic"?	Cognitive and Executive Function Skills			
39. Does the person have time management difficulties?	Cognitive and Executive Function Skills	-	I have good time management skills.	-
40. Does the person have difficulty with large projects, prioritizing and sequencing tasks, setting and keeping to schedules, and knowing when "enough is enough"?	Cognitive and Executive Function Skills	Does the person have difficulty with organization skills?	I have difficulty with organization skills.	I have good organization skills.
41. Does the person have an exceptional long-term memory for events and facts?	Cognitive and Executive Function Skills	Does the person have good long-term memory of facts. Does the person have good long-term memory of events.	I have good long-term memory of facts. I have good long-term memory of events.	- -
42. Does the person report no benefit from meditation, visualization, and similar means of "imaginative" stress management?	Cognitive and Executive Function Skills	-	"Imaginative" stress management such as meditation would benefit me.	-
43. Does the person have intense reaction to change; as much trouble with small changes as big changes? Examples are changes in a route or delivery schedule, restaurant menu changes or moving to a new location.	Cognitive and Executive Function Skills			

44. Does the person frequently engage in black and white thinking? Does he have trouble with "gray areas" and with others "bending the rules"?	Cognitive and Executive Function Skills	Does the person frequently engage in black and white thinking? Does the person have trouble with others "bending the rules"	I believe things are either black or white. "Bending the rules" is unacceptable for me.	- -
45. Does the person have an unusual sense of justice, morality, and notions of proper behavior?	Cognitive and Executive Function Skills			
46. Does the person "tell the truth, and the whole truth" regardless of the circumstances or consequences? Does his frankness get him into trouble?	Cognitive and Executive Function Skills	-	I would tell the truth regardless of the circumstances.	-
47. Is the person drawn to rules and regulations, protocols, procedures, and writing or following directives and standards? Does the person have a high interest in "quality work"?	Cognitive and Executive Function Skills			
48. Is the person a perfectionist? Does he express pleasure with being extraordinarily precise and detail oriented?	Cognitive and Executive Function Skills	Is the person a perfectionist? Does he express pleasure with being extraordinarily detail oriented?	I am a perfectionist. Being detail oriented makes me feel pleasurable.	
49. Does the person find great comfort in performing rote, repetitive tasks that for a person of their intelligence and education is puzzling to others?	Cognitive and Executive Function Skills			
50. Does the person appear to maintain a high state of vigilance and suspicion?	Cognitive and Executive Function Skills	Does the person appear to maintain a high state of suspicion?	I am always suspicious.	-
51. Is the person likely to be intense and alert with matters of interest, and nonchalant or dismissive of matters he doesn't consider important?	Cognitive and Executive Function Skills	Is the person likely to be alert with matters of interest?	I tend to be more alert on things that interest me.	-
52. Does the person have money management difficulties?	Cognitive and Executive Function Skills		I am having issues with money management.	I am having money management issues.
53. Does the person procrastinate, and is he concerned about it?	Cognitive and Executive Function Skills	Does the person procrastinate?	I often find myself procrastinating.	-
54. Is the person subject to unexplainable bouts of impulsivity?	Cognitive and Executive Function Skills			
55. For his age and stage of life, is the person uncharacteristically conservative and equally hesitant about making small and major decisions?	Cognitive and Executive Function Skills	-	I often hesitate in making any forms of decisions.	-
56. For work and other relationships, does the person have difficulty "reading another person's mind"? Examples would be not anticipating and acting to address a work colleague's functional and emotional needs. The same applies to a partner's unarticulated needs for emotional, sexual satisfaction and social as well as physical companionship.	Cognitive and Executive Function Skills			
57. Does the person describe his behavior as being a like a packrat, unable to part with things of little or no intrinsic value? Is the opposite true? Does the person impulsively give things away or discard items known to have a future value?	Cognitive and Executive Function Skills	Does the person describe his behavior as being a like a packrat, unable to part with things of minimal intrinsic value? Does the person impulsively discard	I tend to keep things of minimal intrinsic value.	-

		items known to have a future value?	I always discard items that other find to have a future value.	-
58. Does the person have difficulty in organizing personal records, forgetting appointments and important commitments or constantly misplacing important documents?	Cognitive and Executive Function Skills			
59. Does the person describe himself as clumsy, uncoordinated or prone to accidents? Examples are difficulty performing assembly work, sewing and household repairs, activities requiring bilateral coordination and sensory integration?	Somatic, Motoric and Presentation Issues	-	I am clumsy.	-
60. Does the person not exercise regularly or maintain good physical condition?	Somatic, Motoric and Presentation Issues	Does the person not exercise regularly? Does the person not maintain good physical condition?	- I am maintaining a good physical condition.	I do exercise regularly. I always try to maintain a good physical condition.
61. Does the person have bad feelings about his body and his appearance?	Somatic, Motoric and Presentation Issues	Does the person have bad feelings about his/her appearance?	I am not happy with my appearance.	I am happy with my appearance.
62. Does the person have unusual posture or an unusual walking/running gait?	Somatic, Motoric and Presentation Issues	Does the person have unusual posture?	I have been told that my postures are awkward.	-
63. Even when sitting or in situations that are low-stress, does the person engage in small repetitive, self-stimulatory behaviors? Examples would be knuckle cracking, pencil tapping, fidgeting, hand steepling, grimaces or tics, playing with keys or jewelry, tightening of the jaw, eyebrow arching, scratching, nail-biting, and sighing or low-level vocalization.	Somatic, Motoric and Presentation Issues	-	I have been told that my behaviors are quite repetitive.	-
64. Does the person have digestive difficulties? Examples would be celiac disease, gluten or casein intolerance, and chronic bowel disorders such as irritable bowel syndrome.	Somatic, Motoric and Presentation Issues	-	I always have digestive difficulties.	-
65. Does the person have strong sensory reactions to touch, light, particular sounds, preference for soft or loose clothing, certain odors, texture aversions (extending to certain objects such as paper, and certain foods)? Does the person avoid crowds or have very wide personal space boundaries	Somatic, Motoric and Presentation Issues			
66. Does the person have a marked insensitivity to pain, heat or cold, or have a bad sense of personal safety?	Somatic, Motoric and Presentation Issues			
67. Does the person have bad hygiene and poor self-care habits?	Somatic, Motoric and Presentation Issues			
68. Does the person engage in public self-grooming behavior ordinarily done in private?	Somatic, Motoric and Presentation Issues			
69. Does the person show a preference for a limited range of clothing to the point of his dress being very predictable to others? If “Yes”, would it be accurate to describe this limited wardrobe as a kind of “uniform”?	Somatic, Motoric and Presentation Issues		My choice of clothing is very limited.	
70. Does the person report discomfort cuddling,	Somatic, Motoric and	Does the person report discomfort in	I feel discomfort in hugging.	I feel discomfort when hugging with

hugging, or being held and touched?	Presentation Issues	hugging?		others.
71. Does the person report low sexual desire or a-sexuality? Has the person been repeatedly warned about inappropriate touching or behaviors considered by the object person to be sexual harassment or stalking? Does the person say they never understood dating, or have given up interest in dating?	Somatic, Motoric and Presentation Issues	Does the person report low sexual desire or a-sexuality? Has the person been warned about inappropriate behaviors considered by the object person to be sexual harassment.	I have a low sexual desire. I have been warned about inappropriate behaviors that others considered to be sexual harassment.	- -
72. Does the person have sleep disturbance? Examples would be difficulty falling asleep, waking early, restless sleep and discomfort in sleeping with a partner.	Somatic, Motoric and Presentation Issues	Does the person have sleep disturbance?	I often have problems with my sleep.	-
73. Does the person report difficulty living with others, or becoming independent from parents or other caregivers?	Other Characteristics	Does the person report difficulty living with others?	I do not like to live with others.	-
74. Does the person have elaborate, rigidly-adhered to rituals? Examples are self-care habits, eating, having a work area arranged "just so"," arranging personal effects in a precise order, and ways of getting to and from places.	Other Characteristics			
75. Does the person have a strong attraction to certain visual or auditory patterns?	Other Characteristics	Does the person have a strong attraction to certain visual patterns? Does the person have a strong attraction to certain auditory patterns?	I find certain visual patterns very attractive. Certain auditory patterns attract me very much.	- -
76. Does the person report information about developmental delays or uneven functional development as a child? In school, was he enrolled in a special education program?	Other Characteristics	Does the person report information about development delays? In school, was he enrolled in a special education program?	I have a history of developmental delays. As a child, I was enrolled in a special education program.	- -
77. Does the person consider himself to be emotionally immature? Do you agree?	Other Characteristics	-	I would consider myself as emotionally immature.	-
78. Does the person have any history of seizure activity, or demonstrate absence and other low level seizure behavior?	Other Characteristics	-	I have a history of seizure activity.	-
79. Does the person expend so much energy just getting through the work day or school that he has no energy left for "a life outside" of those activities?	Other Characteristics	Does the person expend so much energy just getting through the workday?	I often spend so much energy just to get through a workday.	-
80. Has the person turned down management positions with statements to the effect that he is not "a management type person"? Has he been promoted to a management position and then demoted or removed due to lack of her people-management skills?	Other Characteristics	Will the person turn down a management position because he is not a management type of person?	I would not pursue a management position because I am not a management type of person.	-
81. Has the person had trouble retaining employment? Is there a long history of many jobs, part-time, unpaid, underpaid work, and temporary or short-duration jobs?	Other Characteristics	Has the person had trouble retaining employment?	Keeping a job is difficult for me.	Maintaining an employment is difficult for me.
Measure:				

Ritvo Autism and Asperger's Diagnostic Scale (RAADS; Ritvo et al., 2008)				
1. I am a sympathetic person.	Social Relatedness	-	-	-
2. I keep many exact words and phrases from movies and television in my memory.	Language and Communication	I keep many exact phrases from movies in my memory.	-	-
3. I understand when friends need to be comforted, and I always try to be helpful.	Social Relatedness	I can easily tell when friends need to be comforted. I always try to be helpful when friends need comfort.	-	-
4. Sometimes I talk too loudly or too softly, and I am not aware of it.	Sensorimotor and Stereotypies	Sometimes I am not aware that I am talking too loud. Sometimes I am not aware that I talking too soft.	-	-
5. I often don't know how to act in social situations.	Social Relatedness	-	-	-
6. I can ''put myself in other people's shoes.''	Social Relatedness	-	-	I can always put myself in others' shoes.
7. I have a hard time understanding the meaning of the phrase: ''He has skeletons in his closet.''	Language and Communication			
8. I don't remember people's faces. I am more likely to remember something about them that others may consider peculiar (like a person's scent).	Language and Communication	I don't remember people's faces. I am more likely to remember people by things that others find odd.	-	I am not good with recognizing people's faces.
9. I would rather tell a ''little white lie'' than hurt someone's feelings.	Social Relatedness	-	-	I would rather tell a white lie than hurting someone's feelings.
10. I always notice how food feels in my mouth. This is just as important to me as how it tastes.	Sensorimotor and Stereotypies	-	-	It is important for me to notice how food feels in my mouth.
11.I miss my best friends or family when we are apart for a long time.	Social Relatedness	I miss my best friends when we are apart for a long time. I miss my family when we are apart for a long time.	-	-
12. Sometimes I offend others by saying what I am thinking. I am not aware that I am doing that, and I am surprised when others tell me that I have been rude.	Social Relatedness	Sometimes I am not aware that I offended others by saying what I was thinking until I was told.	-	-
13. I like to have close friends.	Social Relatedness	-	-	I like having close friends.
14. I'd rather go out to eat in a restaurant by myself than with someone I know.	Social Relatedness		-	I would rather go out to eat alone than with someone I know.
15. I cannot imagine what it would be like to be someone else.	Language and Communication	-	-	-
16. I have been told that I am clumsy and that my posture and gait are awkward.	Sensorimotor and Stereotypies	I have been told that I am clumsy. I have been told that my postures are awkward.	-	-
17. I am very sensitive to the way my clothes feel when I touch them. How they feel is more important to me that how they look.	Sensorimotor and Stereotypies	I am very sensitive to the way my clothes feel. How my clothes feel is more important than how they look.	-	-
18. I like to copy the way certain people speak	Social Relatedness	I like to copy the way certain people	-	-

and act. It helps me appear more normal.		speak. I like to copy the way certain people act.		
19. It can be very intimidating for me to talk to more than one person at the same time.	Social Relatedness	-	-	-
20. I have been told that sometimes I speak too loudly or too softly, even when my voice sounds fine to me.	Sensorimotor and Stereotypies			
21. I have to ‘‘act normal’’ to please other people and make them like me.	Social Relatedness	-	-	I would go with the norm to please other people.
22. Meeting new people is usually easy for me.	Social Relatedness	-	-	-
23. I get highly confused when someone interrupts me when I am talking about something I am very interested in.	Language and Communication	-	-	I get confused when someone interrupts me when I am talking about something I am very interested in.
24. It is difficult for me to understand how other people are feeling when we are talking.	Language and Communication	-	-	-
25. I don’t mind having a conversation with several people at the same time, for instance, around a dinner table, at school, or at work.	Language and Communication	I don’t mind having a conversation with several people at the same time.	-	-
26. I have a hard time figuring out what some phrases mean, like: ‘‘You are the apple of my eye.’’	Language and Communication			
27. It is very difficult for me to understand some emotions (like lust, infatuation, empathy, or embarrassment).	Language and Communication	It is very difficult for me to understand empathy. It is difficult for me to understand embarrassment.	-	-
28. I am more sensitive to smells than anyone I know.	Sensorimotor and Stereotypies	-	-	-
29. Some ordinary textures that do not bother others feel very offensive when they touch my skin.	Sensorimotor and Stereotypies	-	-	Some textures that do not bother others tend to bother me a lot.
30. I get extremely upset when the way I like to do things is suddenly changed.	Sensorimotor and Stereotypies	-	-	I get extremely upset when there is a sudden change of plans.
31. I never wanted or needed to have what other people call an ‘‘intimate relationship.’’	Social Relatedness	-	-	I never wanted to have an intimate relationship.
32. It is difficult for me to start and stop a conversation. I need to keep going until I am finished.	Language and Communication	It is difficult for me to start a conversation. It is difficult for me to end a conversation.	-	It is difficult for me to initiate a conversation.
33. I usually speak in a normal tone.	Sensorimotor and Stereotypies		-	I always speak in a normal tone.
34. I can chat and make small talk with friends and when I meet new people.	Social Relatedness	I can chat with friends comfortably. I can make small talk comfortably with people I just met.	-	-
35. I speak with a normal rhythm and tone.	Sensorimotor and Stereotypies	I speak with a normal rhythm.	-	I usually speak with a normal rhythm.
36. My sensations can suddenly change from very sensitive to very dull.	Sensorimotor and Stereotypies	-	-	
37. When I am shopping, I get very nervous at the	Language and	-	-	When I am shopping, I would worry at

checkout. I have calculated the amount of what I bought in my head, and I worry it will not come out right.	Communication			the checkout that the total amount I calculated will not come out right.
38. Sometimes the sound of a word or a highpitched noise can be painful to my ears.	Sensorimotor and Stereotypies	-	-	Sometimes a high-pitched noise can be painful to my ears.
39. The phrase ''I've got you under my skin'' makes me very uncomfortable.	Language and Communication			
40. I am an understanding type of person.	Social Relatedness	-	-	I am an understanding person.
41. I do not connect with characters in movies and cannot feel what they feel.	Social Relatedness	I do not connect with characters in movies. It is difficult for me to feel for the characters in movies.	-	- I do not usually feel for the characters in movies.
42. I cannot tell when someone is flirting with me.	Social Relatedness	-	-	I can easily tell when someone is flirting with me.
43. I can see in my mind a whole page that I have read, recall an entire long conversation, or remember travel routes in detail, even if these occurred years ago.	Language and Communication	-	-	I can recall details of things that happened years ago.
44. I memorize lists of things that interest me, even when they have no practical use (for example, sports statistics, train schedules, calendar dates, historical facts and dates).	Language and Communication	-	-	I memorize lists of things that interest me, even when they have no practical use.
45. I can tell when someone says one thing but means something else.	Language and Communication	-	-	-
46. I like to talk things over with my friends.	Social Relatedness	-	-	-
47. Sometimes I keep talking and do not notice when others want to say something or are getting bored.	Language and Communication	Sometimes I keep talking and do not notice when others want to say something.	-	-
48. It can be very hard to read someone's face, hand, and body movements when we are talking.	Language and Communication	It can be very hard to read someone's face when we are talking. It can be very hard to read someone's body movement when we are talking.	-	It can be hard to read someone's emotions during a conversation. It can be very hard to read someone's body movement when we are talking.
49. The same thing (like clothes, or temperatures) can feel very different to me at different times.	Sensorimotor and Stereotypies	-	-	The same thing (for example, clothes) can feel very different to me at different times.
50. I feel comfortable with dating and being in social situations with others.	Social Relatedness	I feel comfortable with dating someone. I feel comfortable with being in a social situation with others.	-	-
51. I try to be as helpful as I can when other people tell me their personal problems.	Social Relatedness	-	-	-
52. I have been told that I have an unusual voice (for example, flat, monotone, childish, or high-pitched).	Sensorimotor and Stereotypies	-	-	I have been told that I have an unusual voice.
53. Sometimes a thought or a subject gets stuck in my mind and I have to talk about it even if no one is interested.	Language and Communication	Sometimes if a thought gets stuck in my mind, I have to talk about it even if no one is interested.	-	-
54. I do certain things with my hands over and over again (like flapping, twirling sticks or strings, waving things by my eyes).	Sensorimotor and Stereotypies			

55. I have never been interested in what most of the people I know consider interesting.	Language and Communication	-	-	I have never been interested in what majority of the people consider interesting.
56. I am considered a compassionate type of person.	Social Relatedness	-	-	-
57. I get along with other people by following a set of specific rules that help me to look normal.	Social Relatedness	-	-	-
58. It is almost impossible for me to work and function in groups.	Social Relatedness	It is almost impossible for me to function in groups.	-	-
59. I am considered a loner by those who know me best.	Social Relatedness	-	-	-
60. Sometimes I have to cover my ears to block out painful noises (like vacuum cleaners or people talking too much or too loudly).	Sensorimotor and Stereotypies	-	-	Sometimes I have to cover my ears to block out painful noises (for example, vacuum cleaners).
61. Sometimes things that should feel painful are not (for instance, when I hurt myself or burn my hand on a stove).	Sensorimotor and Stereotypies	-	-	Sometimes things that should feel painful are not (for example, when I hurt myself)
62. Sometimes when I feel overwhelmed by my senses, I have to isolate myself to shut them down.	Sensorimotor and Stereotypies	Sometimes I feel overwhelmed by my senses.	-	-
63. Sometimes when talking to someone, I cannot tell when it is my turn to talk or to listen.	Social Relatedness	Sometimes when talking to someone, I cannot tell when it is my turn to talk.	-	I cannot easily tell when it is my turn to talk during a conversation.
64. When I am talking to someone, it is hard to change the subject. If the other person does so, I can get very upset and confused.	Language and Communication	When I am talking to someone, it is hard to change the subject because I would get upset.	-	When I am talking to someone, it upsets me when the topic is changed.
65. I like things to be exactly the same day after day and even small changes in my routines upset me.	Sensorimotor and Stereotypies	I like things to be exactly the same day after day. Changes to my routine would upset me.	-	-
66. How to make friends and socialize is a mystery to me.	Social Relatedness	-	-	How socialization works is a mystery to me.
67. It calms me to spin around or to rock in a chair when I am feeling stressed.	Sensorimotor and Stereotypies	Spinning around calms me down when I am feeling stressed. Rocking in a chair calms me down when I am feeling stressed.	-	-
68. The phrase, ‘‘Hewears his heart on his sleeve,’’ does not make sense to me.	Language and Communication	-	-	-
69. If I am in a place with many smells, textures to feel, noises, or bright lights; I can get overwhelmed with sensations and feel panicky, anxious, or frightened.	Sensorimotor and Stereotypies	I can easily get overwhelmed with multiple sensations at the same time.	-	-
70. I cannot tell if someone is interested or bored with what I am saying.	Social Relatedness	I cannot tell if someone is interested in what I am saying. I cannot tell if someone is bored with what I am saying.	-	-
71. I like to be by myself as much as I can.	Social Relatedness	-	-	-
72. I keep my thought stacked in my memory like they are on filing cards, and I pick out the ones I need by looking through the stack and finding the right one.	Language and Communication	-	-	-
73. The same sound sometimes seems very loud	Sensorimotor and	-	-	The same sound sometimes seems very

or very soft, even though I know it has not changed.	Stereotypies			different, even though I know it has not changed.
74. I enjoy spending time eating and talking with my family and friends.	Social Relatedness	I enjoy spending time with my family. I enjoy spending time with my friends.	-	-
75. I can't tolerate things I dislike (like smells, textures, sounds, or colors).	Sensorimotor and Stereotypies	-	-	I cannot tolerate things I dislike.
76. I don't like to be hugged and held.	Sensorimotor and Stereotypies	-	-	I don't like to be hugged.
77. When I go somewhere, I have to follow a familiar route or I can get very confused and upset.	Language and Communication	When I go somewhere, I have to follow a familiar route or I can get very upset.	-	-
78. It is difficult to figure out what other people expect of me.	Social Relatedness	-	-	-
Measure: ***The Asperger Syndrome Diagnostic Interview (ASDI; Gillberg et al., 2001)***				
1. Does he/she exhibit considerable difficulties interacting with peers? If so, in what way?	Severe Impairments in Reciprocal Social Interaction (Extreme Egocentricity)	-	It is difficult for me to interact with peers.	-
2. Does he/she exhibit a low degree of concern or a seeming lack of interest in making friends or interacting with peers? If so, please specify:	Severe Impairments in Reciprocal Social Interaction (Extreme Egocentricity)	-	I am not interested in making friends. I am not interested in engaging a conversation with my peers.	-
3. Does he/she have problems appreciating social cues, i.e. does he/she fail to note changes in the social conversation/interaction or to take account of such changes in his/her ongoing interaction with other people? If so, please describe:	Severe Impairments in Reciprocal Social Interaction (Extreme Egocentricity)	-	Social cues are hard for me to understand.	-
4. Does he/she exhibit socially or emotionally inappropriate behaviours? If so, in what way(s)?	Severe Impairments in Reciprocal Social Interaction (Extreme Egocentricity)	Does he/she exhibit socially inappropriate behaviors? Does he/she exhibit emotionally inappropriate behaviors.	I often find myself engaging behaviors that others find inappropriate. I often find myself displaying emotions that are inappropriate to others.	-
5. Is there a pattern of interest or a specific interest which takes up so much of his/her time that time for other activities is clearly restricted? If there is, please comment:	All Absorbing Narrow Interest Pattern(s)	-	I have interest on a specific thing that I often spend so much time on.	-
6. Is there a repetitive quality to his/her interest patterns or specific interest? If so, please specify:	All Absorbing Narrow Interest Pattern(s)			
7. Are his/her interest patterns based more on rote memory than on true meaning?	All Absorbing Narrow Interest Pattern(s)	-	The things that interest me are based more on my routines than on true meaning.	-
8. Does he/she try to introduce and impose routines, rituals or interests on himself/herself in such a way as to produce problems for himself? If so, in what way?	Imposition of Routines, Rituals, and Interests	Does he/she try to introduce routines in such a ways as to produce problems?	I have certain routines that I have to follow or I will feel unease. There are particular rituals I have to follow or I will feel unease.	-
9. Does he/she try to introduce and impose routines, rituals or interests on himself/herself in such a way as to produce problems for others? If so, please describe:	Imposition of Routines, Rituals, and Interests	-	My routines often bother people. My rituals would bother people at times.	-
10. Was his/her language development delayed? If	Speech and Language			

so, please comment:	Peculiarities			
11. Is his/her language 'superficially perfect' regardless of whether or not there are comprehension problems or other speech and language problems? If so, please comment:	Speech and Language Peculiarities			
12. Is his/her language formal, pedantic or 'overly adult'? If so, please describe:	Speech and Language Peculiarities		People often find my language formal.	-
13. Is there any characteristic about his/her voice (pitch, volume, quality, intonation, word stress, 'prosody' etc.) which you find peculiar or unusual? If so, in what way?	Speech and Language Peculiarities			
14. Are there any comprehension problems (including misinterpretations of literal/implied meanings)? If so, what kind of problems?	Speech and Language Peculiarities	-	I have difficulty interpreting the implied meaning of things.	-
15. Does he/she make limited use of gestures? If so, please comment:	Non-verbal Communication Problems			-
16. Is his/her body language awkward, gauche, clumsy, strange or unusual? If so, please comment:	Non-verbal Communication Problems	-	I have been told that my body language is clumsy. I have been told that my body language is strange.	-
17. Are his/her facial expressions limited to a rather small repertoire? If so, please describe:	Non-verbal Communication Problems			
18. Is his/her general expression (including facial) sometimes inappropriate? If so, please describe:	Non-verbal Communication Problems	-	I have been told that my expressions are often inappropriate.	-
19. Is his/her gaze stiff, strange, peculiar, abnormal or odd? If so, please characterize:	Non-verbal Communication Problems	-	I have always been told that my gaze is odd.	-
20. Has he/she been noted to perform poorly on neurodevelopmental examinations either in the past or in connection with the present interview? If so, please comment:	Motor Clumsiness			
Measure: ***Empathy Quotient (Baron-Cohen & Wheelwright, 2003) [Filler items & Items dropped from Factor Analysis were removed]***				
1. I can easily tell if someone else wants to enter a conversation.	Cognitive Empathy	-	-	-
4. I find it difficult to explain to others things that I understand easily, when they don't understand it first time.	Social Skills	-	-	It is difficult for me to explain to people about things that I understand.
6. I really enjoy caring for other people.	Emotional Reactivity	-	-	-
8. I find it hard to know what to do in a social situation.	Social Skills	-	-	-
12. Friendships and relationships are just too difficult, so I tend not to bother with them.	Social Skills	Friendships are just too difficult, so I tend not to bother with them. Relationships are just too difficult, so I tend not to bother with them.	-	I tend not to bother with friendships because they are too difficult. I tend not to bother with relationships because they are too difficult.
14. I often find it difficult to judge if something is rude or polite.	Social Skills	-	-	I find it difficult to judge if someone is being polite.
19. I can pick up quickly if someone says one thing but means another.	Cognitive Empathy			

21. It is hard for me to see why some things upset people so much.	Emotional Reactivity	-	-	-
22. I find it easy to put myself in somebody else's shoes.	Emotional Reactivity			
25. I am good at predicting how someone will feel.	Cognitive Empathy	-	-	-
26. I am quick to spot when someone in a group is feeling awkward or uncomfortable.	Cognitive Empathy	I am quick to spot when someone in a group is feeling uncomfortable.	-	-
27. If I say something that someone else is offended by, I think that that's their problem, not mine.	Emotional Reactivity	-	-	When someone got offended by things that I said, I often think that it was their problem.
29. I can't always see why someone should have felt offended by a remark.	Emotional Reactivity	-	-	-
32. Seeing people cry doesn't really upset me.	Emotional Reactivity	-	-	-
35. I don't tend to find social situations confusing.	Social Skills	-	-	I tend to find social situations confusing.
36. Other people tell me I am good at understanding how they are feeling and what they are thinking.	Cognitive Empathy	-	-	I have been told that I am good at understanding how others are feeling.
41. I can easily tell if someone else is interested or bored with what I am saying.	Cognitive Empathy	I can easily tell if someone is interested with what I am saying. I can easily tell if someone is bored with what I am saying.	-	-
42. I get upset if I see people suffering on news programmes.	Emotional Reactivity	-	-	-
43. Friends usually talk to me about their problems as they say that I am very understanding.	Emotional Reactivity	-	-	-
44. I can sense if I am intruding, even if the other person doesn't tell me.	Cognitive Empathy	-	-	-
48. Other people often say that I am insensitive, though I don't always see why.	Emotional Reactivity	-	-	Other people often say that I am insensitive, though I don't see why.
50. I usually stay emotionally detached when watching a film.	Emotional Reactivity	-	-	-
52. I can tune into how someone else feels rapidly and intuitively.	Cognitive Empathy	-	-	I can tune into how someone else feels intuitively.
54. I can easily work out what another person might want to talk about.	Cognitive Empathy	-	-	-
55. I can tell if someone is masking their true emotion.	Cognitive Empathy	-	-	I can easily tell if someone is masking their true emotion.
57. I don't consciously work out the rules of social situations.	Social Skills	-	-	I would consciously work out the rules of social situations.
58. I am good at predicting what someone will do.	Cognitive Empathy	-	-	-
59. I tend to get emotionally involved with a friend's problems.	Emotional Reactivity	-	-	-
Measure: ***Autism-Spectrum Quotient – Adult Version (Baron-Cohen et al., 2001)***				
1. I prefer to do things with others rather than on my own.	N/A		-	
2. I prefer to do things the same way over and	N/A	-	-	-

over *again*.				
3. If I try to imagine something, I find it very easy to create a picture in my mind.	N/A	-	-	-
4. I frequently get so strongly absorbed in one thing that I lose sight of other things.	N/A	-	-	-
5. I often notice small sounds when others do not.	N/A	-	-	-
6. I usually notice car number plates or similar strings of information.	N/A	-	-	I usually notice car number plates.
7. Other people frequently tell me that what I've said is impolite, even though I think it is polite.	N/A	-	-	-
8. When I'm reading a story, I can easily imagine what the characters might look like.	N/A	-	-	-
9. I am fascinated by dates.	N/A	-	-	-
10. In a social group, I can easily keep track of several different people's conversations.	N/A	-	-	-
11. I find social situations easy.	N/A	-	-	-
12. I tend to notice details that others do not.	N/A	-	-	-
13. I would rather go to a library than a party.	N/A			
14. I find making up stories easy.	N/A	-	-	-
15. I find myself drawn more strongly to people than to things.	N/A	-	-	-
16. I tend to have very strong interests, which I get upset about if I can't pursue.	N/A	-	-	-
17. I enjoy social chit-chat.	N/A	-	-	-
18. When I talk, it isn't always easy for others to get a word in edgeways.	N/A			
19. I am fascinated by numbers.	N/A	-	-	-
20. When I'm reading a story, I find it difficult to work out the characters' intentions.	N/A	-	-	-
21. I don't particularly enjoy reading fiction.	N/A	-	-	-
22. I find it hard to make new friends.	N/A	-	-	-
23. I notice patterns in things all the time.	N/A	-	-	-
24. I would rather go to the theatre than a museum.	N/A	-	-	-
25. It does not upset me if my daily routine is disturbed.	N/A	-	-	It upsets me if my daily routine is disturbed.
26. I frequently find that I don't know how to keep a conversation going.	N/A	-	-	-
27. I find it easy to "read between the lines" when someone is talking to me.	N/A	-	-	-
28. I usually concentrate more on the whole picture, rather than the small details.	N/A	-	-	-
29. I am not very good at remembering phone numbers.	N/A	-	-	-
30. I don't usually notice small changes in a situation, or a person's appearance.	N/A	I usually don't notice small changes in a situation.	-	I tend to notice small changes in a situation.
31. I know how to tell if someone listening to me is getting bored.	N/A	-	-	-
32. I find it easy to do more than one thing at once.	N/A	-	-	-
33. When I talk on the phone, I'm not sure when it's my turn to speak.	N/A	-	-	-

34. I enjoy doing things spontaneously.	N/A	-	-	-
35. I am often the last to understand the point of a joke.	N/A	-	-	-
36. I find it easy to work out what someone is thinking or feeling just by looking at their face.	N/A	I find it easy to work out what someone is feeling by looking at their face.	-	-
37. If there is an interruption, I can switch back to what I was doing very quickly.	N/A	-	-	-
38. I am good at social chit-chat.	N/A	-	-	-
39. People often tell me that I keep going on and on about the same thing.	N/A	-	-	-
40. When I was young, I used to enjoy playing games involving pretending with other children.	N/A	-	-	When I was young, I used to enjoy playing games with other children that involved pretending.
41. I like to collect information about categories of things (e.g., types of car, types of bird, types of train, types of plant, etc.).	N/A	-	-	-
42. I find it difficult to imagine what it would be like to be someone else.	N/A	-	-	-
43. I like to plan any activities I participate in carefully.	N/A	-	-	-
44. I enjoy social occasions.	N/A	-	-	-
45. I find it difficult to work out people's intentions.	N/A	-	-	-
46. New situations make me anxious.	N/A	-	-	-
47. I enjoy meeting new people.	N/A	-	-	-
48. I am a good diplomat.	N/A	-	-	-
49. I am not very good at remembering people's date of birth.	N/A	-	-	-
50. I find it very easy to play games with children that involve pretending.	N/A	-	-	-

Appendix E Item Content Rating Review Forms

Empathy

Category		**Category Specifications**							
Empathy		**Empathy** is the ability to understand and put oneself into others' feelings (Baron-Cohen & Wheelwright, 2004). It is one's reaction upon observing others' experiences (Davis, 1980).							
		Item Rating Scale: 0 "No Fit" 1 2 3 4 "Excellent Fit"							
Item #	Item	Item Rating					Suggested Item Revisions	Is the Item Well Written?	Is the Item Essential to the Domain?
1.	It is difficult for me to understand others' feelings.	0	1	2	3	4		Yes / No	Yes / No
2.	I expect other people to know what I have experienced.	0	1	2	3	4		Yes / No	Yes / No
3.	I am a sympathetic person.	0	1	2	3	4		Yes / No	Yes / No
4.	I can easily tell when friends need to be comforted.	0	1	2	3	4		Yes / No	Yes / No
5.	I can always put myself in others' shoes.	0	1	2	3	4		Yes / No	Yes / No
6.	I cannot imagine what it would be like to be someone else.	0	1	2	3	4		Yes / No	Yes / No
7.	I don't know how to respond when other people are embarrassed.	0	1	2	3	4		Yes / No	Yes / No
8.	I am an understanding person when people tell me their problems.	0	1	2	3	4		Yes / No	Yes / No
9.	I do not usually feel for the characters in movies.	0	1	2	3	4		Yes / No	Yes / No
10.	I enjoy reading the newspaper.	0	1	2	3	4		Yes / No	Yes / No
11.	I am good at predicting how someone will feel.	0	1	2	3	4		Yes / No	Yes / No
12.	I am quick to spot when someone in a group is feeling uncomfortable.	0	1	2	3	4		Yes / No	Yes / No
13.	I have been told that I am good at understanding how others are feeling.	0	1	2	3	4		Yes / No	Yes / No

Category Empathy		**Category Specifications** **Empathy** is the ability to understand and put oneself into others' feelings (Baron-Cohen & Wheelwright, 2004). It is one's reaction upon observing others' experiences (Davis, 1980).							
14.	Other people often say that I am insensitive, though I don't see why.	0	1	2	3	4		Yes / No	Yes / No
15.	I can tell if someone is masking their true emotions.	0	1	2	3	4		Yes / No	Yes / No
16.	I would rather tell a white lie than hurt someone's feelings.	0	1	2	3	4		Yes / No	Yes / No
17.	I am considered a compassionate person.	0	1	2	3	4		Yes / No	Yes / No

If you were to group together the items you rated at 3 or 4, would these items completely represent the category of Empathy?

YES **NO**

If **NO**, please indicate what items should be added:

Additional Comments:

Stereotyped and Restricted Repetitive Patterns of Interests and Activities

Category Stereotyped and Restricted Repetitive Patterns of Interests and Activities		**Category Specifications** **Stereotyped and Restricted Repetitive Patterns of Interests and Activities** is referring to an intense interest on certain things and/or activities. This also includes strictly following a routine by an individual (APA, 2000).							
		Item Rating Scale: 0 1 2 3 4 "No Fit" "Excellent Fit"							
Item #	Item	Item Rating					Suggested Item Revisions	Is the Item Well Written?	Is the Item Essential to the Domain?
1.	I feel distressed when things do not go as expected.	0	1	2	3	4		Yes / No	Yes / No
2.	I have difficulty multitasking.	0	1	2	3	4		Yes / No	Yes / No
3.	Being detail oriented is pleasurable to me.	0	1	2	3	4		Yes / No	Yes / No
4.	I tend to be more alert on things that interest me.	0	1	2	3	4		Yes / No	Yes / No
5.	My choice of clothing is very limited.	0	1	2	3	4		Yes / No	Yes / No
6.	I find certain visual patterns very attractive (e.g., the arrangement of objects, the arrangement of colours)	0	1	2	3	4		Yes / No	Yes / No
7.	I memorize lists of things that interest me, even when they have no practical use.	0	1	2	3	4		Yes / No	Yes / No
8.	When I go somewhere, I have to follow a familiar route or I can get very upset.	0	1	2	3	4		Yes / No	Yes / No
9.	I like things to be exactly the same day after day.	0	1	2	3	4		Yes / No	Yes / No
10.	I have certain routines that I have to follow or I will feel unease.	0	1	2	3	4		Yes / No	Yes / No

Category Stereotyped and Restricted Repetitive Patterns of Interests and Activities		**Category Specifications** **Stereotyped and Restricted Repetitive Patterns of Interests and Activities** is referring to an intense interest on certain things and/o4r activities. This also includes strictly following a routine by an individual (APA, 2000).							
11.	My routines often bother people.	0	1	2	3	4		Yes / No	Yes / No
12.	Changes to my routine would upset me.	0	1	2	3	4		Yes / No	Yes / No
13.	I tend to notice details that others do not.	0	1	2	3	4		Yes / No	Yes / No
14.	I prefer take-out food over dining in.	0	1	2	3	4		Yes / No	Yes / No
15.	I tend to have very strong interests, which I get upset about if I can't pursue.	0	1	2	3	4		Yes / No	Yes / No
16.	I am fascinated by numbers.	0	1	2	3	4		Yes / No	Yes / No
17.	I notice patterns in things all the time.	0	1	2	3	4		Yes / No	Yes / No
18.	I tend to notice small changes in a situation.	0	1	2	3	4		Yes / No	Yes / No
19.	I get extremely upset when there is a sudden change of plans.	0	1	2	3	4		Yes / No	Yes / No

If you were to group together the items you rated at 3 or 4, would these items completely represent the category of Stereotyped and Restricted Repetitive Patterns of Interests and Activities?

YES **NO**

If **NO**, please indicate what items should be added:

Additional Comments:

Deficit in Social Communication Skills: Other Problems in Communication Skills

Category Deficit in Social Communication Skills: Other Problems in Communication Skills		**Category Specifications** ***Other Problems in Communication Skills*** include the displaying of inappropriate motor actions by an individual during social interactions, personal feelings of inability to connect with others, inability to understand strategies required in a social situation, and negative preference on reacting to social situations.							
		Item Rating Scale: 0 1 2 3 4 "No Fit" "Excellent Fit"							
Item #	Item	Item Rating					Suggested Item Revisions	Is the Item Well Written?	Is the Item Essential to the Domain?
1.	I am concerned about my loneliness.	0	1	2	3	4		Yes / No	Yes / No
2.	I prefer to be alone than in a group.	0	1	2	3	4		Yes / No	Yes / No
3.	I am frustrated about not having friends.	0	1	2	3	4		Yes / No	Yes / No
4.	I always think that my choice is the only choice that others should go with.	0	1	2	3	4		Yes / No	Yes / No
5.	I enjoy working on puzzles.	0	1	2	3	4		Yes / No	Yes / No
6.	I would not easily change my behaviour even if others were present.	0	1	2	3	4		Yes / No	Yes / No
7.	It is difficult for me to accept criticism.	0	1	2	3	4		Yes / No	Yes / No
8.	I find conflict management difficult.	0	1	2	3	4		Yes / No	Yes / No
9.	I understand why people need personal space.	0	1	2	3	4		Yes / No	Yes / No
10.	I understand the reason for others' need of privacy.	0	1	2	3	4		Yes / No	Yes / No
11.	I would rather go out alone than with someone I know.	0	1	2	3	4		Yes / No	Yes / No
12.	Meeting new people is usually easy for me.	0	1	2	3	4		Yes / No	Yes / No

Category Deficit in Social Communication Skills: Other Problems in Communication Skills		**Category Specifications** ***Other Problems in Communication Skills*** include the displaying of inappropriate motor actions by an individual during social interactions, personal feelings of inability to connect with others, inability to understand strategies required in a social situation, and negative preference on reacting to social situations.							
13.	I feel comfortable being in a social situation with others.	0	1	2	3	4		Yes / No	Yes / No
14.	It is almost impossible for me to function in groups.	0	1	2	3	4		Yes / No	Yes / No
15.	I am considered a loner by those who know me best.	0	1	2	3	4		Yes / No	Yes / No
16.	It is difficult to figure out what other people expect of me.	0	1	2	3	4		Yes / No	Yes / No
17.	It is hard for me to see why some things upset people so much.	0	1	2	3	4		Yes / No	Yes / No
18.	New situations make me anxious.	0	1	2	3	4		Yes / No	Yes / No
19.	I enjoy meeting new people.	0	1	2	3	4		Yes / No	Yes / No

If you were to group together the items you rated at 3 or 4, would these items completely represent the category of Other Problems in Social Communication Skills?

YES **NO**

If **NO**, please indicate what items should be added:

Additional Comments:

Deficit in Social Communication Skills: Problems in Pragmatic Language

Category Deficit in Social Communication Skills: Problems in Pragmatic Language		**Category Specifications** ***Problems in Pragmatic Language*** refers to difficulties in social use of language -understanding and applying the rules in social communication. This involves the lack of abilities with respect to: use of language, change of language, and following rules of communication (American Speech-Language-Hearing Association, 2009). Examples include the use of facial expressions, verbal and non-verbal signals.							
		Item Rating Scale: 0 1 2 3 4 "No Fit" "Excellent Fit"							
Item #	Item	Item Rating					Suggested Item Revisions	Is the Item Well Written?	Is the Item Essential to the Domain?
1.	I have been told that my words are often unrelated to the conversation.	0	1	2	3	4		Yes / No	Yes / No
2.	I am often told not to interrupt when others are talking.	0	1	2	3	4		Yes / No	Yes / No
3.	I am often told that I should look at the people I am talking to.	0	1	2	3	4		Yes / No	Yes / No
4.	Sometimes I am not aware that my words have made people feel uncomfortable.	0	1	2	3	4		Yes / No	Yes / No
5.	I don't mind having a conversation with several people at the same time.	0	1	2	3	4		Yes / No	Yes / No
6.	It is difficult for me to initiate conversation.	0	1	2	3	4		Yes / No	Yes / No
7.	I am good at making friendly conversation with people I just met.	0	1	2	3	4		Yes / No	Yes / No
8.	I find it difficult to know when it is my turn to talk during a conversation.	0	1	2	3	4		Yes / No	Yes / No
9.	I can easily tell if someone is interested in what I am saying.	0	1	2	3	4		Yes / No	Yes / No
10.	It is difficult for me to understand social cues (e.g., body language).	0	1	2	3	4		Yes / No	Yes / No

Category Deficit in Social Communication Skills: Problems in Pragmatic Language		**Category Specifications** ***Problems in Pragmatic Language*** refers to difficulties in social use of language -understanding and applying the rules in social communication. This involves the lack of abilities with respect to: use of language, change of language, and following rules of communication (American Speech-Language-Hearing Association, 2009). Examples include the use of facial expressions, verbal and non-verbal signals.							
11.	I have been told that my facial expressions are often inappropriate.	0	1	2	3	4		Yes / No	Yes / No
12.	I find it difficult to know when someone is being polite.	0	1	2	3	4		Yes / No	Yes / No
13.	I have difficulty knowing how to keep a conversation going.	0	1	2	3	4		Yes / No	Yes / No
14.	I am often the last to understand the point of a joke.	0	1	2	3	4		Yes / No	Yes / No
15.	I often have difficulty ending a conversation.	0	1	2	3	4		Yes / No	Yes / No
16.	People often find my language formal.	0	1	2	3	4		Yes / No	Yes / No
17.	When I take tests, I tend to do better on multiple-choice questions than essay questions.	0	1	2	3	4		Yes / No	Yes / No
18.	Sometimes I am not aware that I am talking too loudly.	0	1	2	3	4		Yes / No	Yes / No
19.	I have been told that my tone of voice does not change when I speak.	0	1	2	3	4		Yes / No	Yes / No

If you were to group together the items you rated at 3 or 4, would these items completely represent the category of Problems in Pragmatic Language?

YES **NO**

If **NO**, please indicate what items should be added:

Additional Comments:

Sensorimotor and Stereotyped Patterns of Behaviours

Category Sensorimotor and Stereotyped Patterns of Behaviours		**Category Specifications** **Sensorimotor and Setereotyped Patterns of Behaviours** is defined as displaying of unusual motor actions by an individual regularly (APA, 2000). Hyper- and/or hypo- sensitivity on objects/things (Barnhill, 2007; Hurlbutt & Chalmers, 2004) are also included in this category.							
		Item Rating Scale: 0 1 2 3 4 "No Fit" "Excellent Fit"							
Item #	Item	Item Rating					Suggested Item Revisions	Is the Item Well Written?	Is the Item Essential to the Domain?
1.	I have been told that my postures are awkward	0	1	2	3	4		Yes / No	Yes / No
2.	I have been told that my behaviours are repetitive.	0	1	2	3	4		Yes / No	Yes / No
3.	I enjoy having my picture taken.	0	1	2	3	4		Yes / No	Yes / No
4.	I avoid hugging others.	0	1	2	3	4		Yes / No	Yes / No
5.	It is important for me to notice how food feels in my mouth.	0	1	2	3	4		Yes / No	Yes / No
6.	I am very sensitive to the way my clothes feel on my skin.	0	1	2	3	4		Yes / No	Yes / No
7.	How my clothes feel is more important than how they look.	0	1	2	3	4		Yes / No	Yes / No
8.	I am sensitive to smells.	0	1	2	3	4		Yes / No	Yes / No
9.	Some textures that do not bother others tend to bother me a lot.	0	1	2	3	4		Yes / No	Yes / No
10.	Sometimes a high-pitched noise can be painful to my ears.	0	1	2	3	4		Yes / No	Yes / No
11.	I have been told that I have an unusual voice.	0	1	2	3	4		Yes / No	Yes / No
12.	Sometimes I have to cover my ears to block out painful noises (for example, vacuum cleaners).	0	1	2	3	4		Yes / No	Yes / No

Category Sensorimotor and Stereotyped Patterns of Behaviours		**Category Specifications** **Sensorimotor and Setereotyped Patterns of Behaviours** is defined as displaying of unusual motor actions by an individual regularly (APA, 2000). Hyper- and/or hypo- sensitivity on objects/things (Barnhill, 2007; Hurlbutt & Chalmers, 2004) are also included in this category.							
13.	Sometimes things that should feel painful are not (for example, when I hurt myself).	0	1	2	3	4		Yes / No	Yes / No
14.	I often notice small sounds when others do not.	0	1	2	3	4		Yes / No	Yes / No
15.	Spinning around calms me down when I am feeling stressed.	0	1	2	3	4		Yes / No	Yes / No
16.	I can easily get overwhelmed with multiple sensations at the same time.	0	1	2	3	4		Yes / No	Yes / No
17.	I often find myself engaging in behaviours that others find inappropriate.	0	1	2	3	4		Yes / No	Yes / No

If you were to group together the items you rated at 3 or 4, would these items completely represent the category of Sensorimotor and Stereotyped Patterns of Behaviours?

YES **NO**

If **NO**, please indicate what items should be added:

Additional Comments