Spring 5-7-2010

Investigating the functionality of a self-report instrument to detect autistic traits in a non-clinical college population: Psychometric properties of the short version of autism-spectrum quotient (AQ-26)

Anna Zilberberg
James Madison University

Investigating the Functionality of a Self-Report Instrument to Detect Autistic Traits in a Non-Clinical College Population: Psychometric Properties of the Short Version of Autism-Spectrum Quotient (AQ-26)

Anna Zilberberg

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts

Department of Graduate Psychology

May 2010

Acknowledgments

This thesis project would never be possible without those who helped me along the way. I am especially grateful to the individuals who served on my thesis committee. First and foremost, I would like to acknowledge the invaluable contribution that my advisor, Dr. Dena A. Pastor made to the current work. She went beyond her duty and devoted extra time and effort to this project. I will always be striving for the level of diligence, conscientiousness, and overall superb work ethic of Dr. Pastor. I also owe special thanks to Dr. J. Christine Harmes for her prompt feedback and readiness to help, even when she was out of town. I am also very grateful to Brenda Fogus and Dr. Trevor Stokes for providing their expert opinion on the subject matter of this work. In addition, I would like to thank Dr. Sara J. Finney for her advice on the technical issues I encountered along the way.

In addition to JMU affiliates who made this work possible, there are others whose contributions should be recognized. In particular, I would like to thank Dr. Diane Zager, a researcher specializing in autism spectrum disorders at Pace University, NY, for meeting me and sharing her expertise at the regional conference in the fall 2009. Also, I would like to express my deepest gratitude to Tim Dempsey, a graduate student at Louisiana State University who conducts research in the area of assessment and screening of autistic disorders. Tim has become my pen pal over the course of the last year with whom I openly shared ideas and concerns. I would also like to thank Megan France and Daniel Jurich for encouraging me along the way. Lastly, I would like to thank all of my friends and family for their continuous and unconditional support throughout this journey. I would never be able to finish this work without everyone's help.

Table of Contents

List of Tables

List of Figures

Abstract

The present study investigated the dimensionality of the short version of Autism-Spectrum Quotient (AQ-26) (Baron-Cohen et al., 2001) via confirmatory factor analysis (CFA) and exploratory factor analysis (EFA). Designed to screen for autistic traits in a non-clinical adult population, the AQ-26 can potentially be a very useful tool both in research and practice. However, evidence pertaining to the structural validity of the AQ-26 is scarce and inconclusive. Competing factor structure models based on previous research were specified and tested using an American college student sample. None of the theoretically specified models provided adequate fit for the data and the focus of the analysis switched to exploring alternative models and analyzing misfit. Although the structural validity of the AQ-26 was not supported, suggestions for future instrument revisions were made based on the results. Additionally, two scoring schemes were deemed not interchangeable, and the implications of using them were discussed. In summary, the analyses indicated that the AQ-26 needs substantial revision before it can be used in research or practice.

CHAPTER 1

Introduction

Asperger Syndrome

Asperger Syndrome (AS) is a life-long developmental condition occupying a higher functioning end of the autism spectrum. It is not characterized by cognitive and language delays, but is typified by the autistic triad of impairments: social skills deficiencies, repetitive behaviors, and communication difficulties, as well as other features (Myles & Simpson, 2002). According to the Center for Disease Control and Prevention (2007), the current prevalence rate of autistic disorders (including AS) in the United States of America, is 1 in 150. Although impairments associated with AS are not as debilitating as those characteristic of more extreme autistic disorders, they nonetheless present significant barriers to optimal functioning of those affected. In fact, the mild nature of AS symptoms and lack of cognitive delay render this condition an invisible disability – likely to be common in the general population but mistaken for mere eccentricity (Attwood, 2007). Current understanding of the AS traits among adults is especially limited, due to the lack of large-scale research studies with the adult population (Tantam, 2000). This is unsettling given that there is reason to suspect that young adults pursuing college education and exhibiting mild autistic behaviors face unique challenges in both the academic and social domains of college life (Glennon, 2001; Smith, 2007). Lack of assessment instruments suitable for screening for autistic traits in the student population likely contributes to the poor understanding of this problem.

Measurement of Asperger Syndrome

Measures suitable for screening for AS traits in the general student population are few, and warrant further study. Although there is a multitude of tools used to diagnose autism in clinical practice, none of these tools is appropriate for the purposes of screening for AS traits among students. These instruments are not a good match for the purposes of screening adults because most of them require one-on-one administration by a trained clinician or an interview with an informant (e.g., Myles, Bock, & Simpson, 2001), while others are designed exclusively for children (Gilliam, 2001). It is not feasible to use such instruments for screening of the general student population because such instruments require qualified personnel and third-party involvement. In her review of assessment tools for AS, Howlin (2000) emphasizes the paucity of AS measures in spite of "the pressing need to identify those on the autism spectrum whose deficits are subtle yet still have a major impact on their lives" (p. 127). Howlin concluded that developing assessment instruments for the AS is premature due to the lack of consensus on diagnostic criteria for AS. However, several instruments designed to assess AS have been published since 2000, indicating that a collaborative effort to refine the diagnostic criteria has been launched. In the most recent review of the data-based methods for assessing AS, Matson and Boisjoli (2008) examined a number of clinical tools for assessing AS and found downfalls with all of them. Nonetheless, Matson and Boisjoli recognized the emergence of such instruments as a good harbinger – signifying that "research teams are 'in the hunt' to develop reliable and valid diagnostic instruments" (p. 245). Such

screening instruments are especially needed for investigating AS traits among college students, who represent an under-served population in this regard.

In order to better understand the distribution and prevalence of autistic traits among college students, a sound assessment tool is needed that is confidential, efficient, and non-intrusive. One such self-report screening measure, the Autism-Spectrum Quotient (AQ), was recently created by a research team in London (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001)[1]. Designed specifically to assess autistic traits in the adult non-clinical population, the AQ has since been used in research examining AS traits in different countries (e.g., Voracek & Dressler, 2006), relationships of AS traits with the big five personality traits (e.g., Austin, 2005), schizotypal personality traits (Hurst, Nelson-Gray, Mitchell, & Kwapil, 2006), differences between scientists and non-scientists (Hoekstra, Bartels, Cath, & Boomsma, 2008) and sensitivity (Liss, Maillopux, & Erchull, 2008). The original instrument consists of 50 items and is most commonly used in research (for clarity, the full 50 item version of the AQ will be hereafter referred to as the AQ-50). However, Austin (2005) identified 26 items that appeared to be most salient to the scale components and presented them in her study, prompting other researchers to consider these items in isolation (this scale will be hereafter referred to as the AQ-26). Although the research investigating the relationships between the two AQ scales and other variables can provide fruitful validity information,

[1] Two other self-report measures tapping into AS traits have recently been published: Adult Asperger Syndrome Scale (AASS) (Foster, 2002) and Ritvo Autism and Asperger Diagnostic Scale (RAADS) (Ritvo, Ritvo, Guthrie, Yuwiler, Ritvo, & Weisbender, 2008). Even though these two instruments hold great promise, they have not yet been studied by anybody but their authors. Due to the lack of empirical evidence for these two instruments, they will not be considered in great detail in this study.

this line of research might be somewhat premature, given that the dimensionality of both scales is still unclear. Specifically, the AQ-26 lacks empirical support despite its potential as a more parsimonious measure. For this reason, the current study focused on the AQ-26, while taking previous research on both the AQ-50 and the AQ-26 into account (The AQ-26 can be found in the Appendix A).

Validity Evidence for the AQ-50 and the AQ-26

According to the *Standards for Educational and Psychological Testing* (1999), validity of an instrument refers to "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (p.9). Importantly, validity is not an all or nothing property of a test, but rather a body of empirical evidence supporting the interpretation of test *scores.* In particular, construct validation refers to the cyclical, multi-stage process of determining whether the instrument measures the construct it purports to measure. As such, it is imperative to gather construct validity evidence for an instrument before using it in research or practice.

In accordance with Benson's (1998) strong program of construct validation, the following stages of construct validation, as they pertain to the AQ, will be examined in the next sections: (a) substantive stage, subsuming theoretical and empirical domains underlying the scale, and (b) structural stage, subsuming internal structure of the scale. The third stage in the Benson's framework involves evaluating the external validity of the scale, or the relationships with related constructs and the scale's capability to differentiate between groups. However, the external stage will not be examined in the current study because the structural stage needs substantial work and must be thoroughly addressed

first. Collecting empirical evidence for the structural validity of the instrument *prior* to examining how its scores relate to other variables is very important. Psychometric studies can shed light on the dimensionality of the scale and suggest the correct scoring procedure (total score versus subscales). In addition, research focusing on the internal domain of the AQ can provide information about the reliability of the subscales and quality of the items. Before proceeding to discussing the psychometric studies of the scale, it is necessary to consider the substantive validity evidence and the scale development process.

Development of the AQ-50

The first of its kind, the AQ-50 was created by a team of researchers and practitioners at the Autism Research Center in London, Great Britain in 2001. Baron-Cohen and his colleagues recognized the need for a brief self-report measure to screen for autistic traits and developed a 50-item Likert-type measure using current scientific understanding of AS as well as their own clinical expertise. The use of self-report was considered appropriate because individuals with AS traits do not suffer from a cognitive impairment and are thus fully capable of responding accurately. Initial scale development underwent rigorous procedures, including comprehension checks, piloting the instrument on adults with and without disabilities, and item revision and deletion. In general, the authors were guided by the theoretical assumption that autistic traits "lie on a *continuum* of social-communication disability, with AS as the bridge between autism and normality" (Baron-Cohen et al., 2001, p. 6, emphasis in the original) and it is therefore reasonable to create a quantitative measure. Further, they worded some of the items in terms of

preference rather than ability in case some individuals are incapable of accurately judging their own abilities. Approximately half of the items were negatively worded. The test developers recommended calculating the total AQ-50 score by following this procedure: response categories 1 (definitely agree) and 2 (agree) are scored as 1 and response categories 3 (disagree) and 4 (definitely disagree) are scored as 0, resulting in the highest possible score of 50, with higher values purportedly indicating heavier autistic load. Such recommendation prompted others to use and interpret the total AQ-50 score, although some kept the four-point scoring scheme (e.g. Stewart & Austin, 2005). After the publication of AQ-50, several researchers investigated the structural properties of this measure, yielding competing models regarding the dimensionality of the scale. The following section outlines these models.

Dimensionality of the AQ

Investigating the dimensionality underlying responses to the AQ is imperative to gathering structural validity evidence for the scale. Further, understanding the factor structure of the measure is necessary for scoring the measure appropriately. For example, if a single latent factor drives the responses to all of the items, then a total score can be meaningfully interpreted. Similarly, if five distinct factors underlie the scale, then five subscale scores are appropriate. Recognizing the need to explore the structure of the AQ, several researchers proposed competing models. These models are presented below in the chronological order in which they appeared in the literature.

1. Single-Factor (*Figure 1)*
2. Five-Factor (Baron-Cohen et al., 2001) (*Figure 2)*

3. Three-Factor (Austin, 2005; Hurst et al., 2007) (*Figure 3)*
4. Higher-Order Factor (Hoekstra et al., 2008) (*Figure 4)*
5. Four-Factor (Stewart & Austin, 2009) (*Figure 5*)

The detailed description of each one of these studies can be found in Chapter 2 of this work. Some of the models specified above are theoretically-based, others are empirically derived, and others are implied from the use of the scale. The first, single-factor model is implied based on the use of the scale as well as the authors' recommendation. Baron-Cohen et al. (2001) recommended calculating a total score, and some researchers followed their lead. However, such a score is only meaningful if a single latent factor underlies the responses to the AQ-26. Research thus far indicates that a single-factor model is not a plausible solution (Hoekstra et al., 2008). However, the total score is still being used in research (e.g., Stewart & Austin, 2009). Thus, further investigation is needed to test the plausibility of a single-factor model.

The five-factor model is specified based on the authors' conceptualization of AS. Although this model has strong theoretical groundings, it has not garnered much support in the empirical studies. For example, Hurst et al. (2007) and Hoekstra et al. (2008) rejected this model. Nonetheless, the five-dimensional conceptualization of AS warrants further study as the original theoretical model.

The three-factor model is rather strong because it is supported by two empirical studies (Austin, 2005; Hurst et al., 2007) and also has some theoretical support because it better aligns with the autistic triad of impairments. Notably, both of these studies used principal component analysis (PCA) which does not account for measurement error and

thus fails to uncover the latent factors (Preacher & MacCallum, 2003). Moreover, these two studies used samples from different countries and scored the measure differently, further limiting the comparability between them.

The higher-order factor championed by Hoekstra et al. (2008) resulted from a confirmatory factor analysis (CFA). This study was by far the strongest because it used confirmatory techniques to test previously specified models and accounted for the categorical nature of the data. However, the fit indices reported by the authors did not reach the recommended cutoff values, making the reader suspect of the model-data fit. Unfortunately, the three-factor model was not tested by Hoekstra et al. (2008).

The four-factor model championed by Stewart and Austin (2005) resulted from the analysis in which confirmatory techniques were used for exploratory purposes. In other words, the authors derived CFA indices for various exploratory models. Although the results of such analyses are not as compelling as those of traditional CFA, they nonetheless contribute to our understanding of the AQ dimensionality.

Limitations

It is important to note that most of the models outlined above were championed for the AQ-50, not the AQ-26. Nonetheless, these structural studies inform our hypotheses about the dimensionality of the AQ-26. In the current study, the models specified a priori to underlie the AQ-26 are based on the previous research conducted on the AQ-50. Although these studies are very informative, they have some limitations that need to be noted. The following section briefly outlines these limitations and implications for the current study.

The structural studies described above have been conducted using samples from different countries, limiting the generalizability of these results to an American sample. Specifically, three studies used British student samples (Austin, 2005; Baron-Cohen et al., 2001; Stewart & Austin, 2009), one study was conducted using a Dutch sample (Hoekstra et al., 2008), and only a single study used an American sample (Hurst et al., 2007). Unfortunately, measurement invariance of the AQ has not been investigated; therefore, we cannot say whether the scale functions equivalently across nations (Steenkamp & Baumgartner, 1998).

Inconsistency in the use of scoring schemes exacerbates the issue of generalizing the results across studies. Recall that there are two different scoring schemes available for the AQ: two-point dichotomous and four-point. Two of the structural studies reviewed used the dichotomous scoring: the first publication by Baron-Cohen et al. (2001) and the study conducted by Hurst et al. (2007). The other studies used the four-point scoring scheme. Our lack of knowledge regarding the comparability of these two scoring schemes limits the conclusions we can make from these structural studies.

Another issue with these studies pertains to the categorical nature of the variables. Whether scored on a two-point or a four-point scale, the variables are categorical and thus cannot be normally distributed by definition, or linearly related to factors (Finney & DiStefano, 2006). Therefore, factor analytic methods with the assumption of normal distribution need to be used in order to achieve accurate results. However, only a single study (Hoekstra et al., 2008) accounted for this.

Yet another issue pertains to the use of different factor analytic techniques across these studies. The first study by Baron-Cohen et al. (2001) did not factor analyze the scale at all. Two studies (Austin, 2005; Hurst et al., 2007) used PCAs that are exploratory in nature and do not account for measurement error (Preacher & MacCallum, 2003). Only one research team (Hoekstra et al., 2008) used CFA with the appropriate estimation method for the categorical data. Unfortunately, the latter study has important limitations: only a few of the a priori models were tested, poorly functioning fit indices were interpreted, and model misfit was not diagnosed. These inconsistencies and limitations among structural studies lead us to the formulation of the current research study.

Purpose of the Current Study

Overall, the evidence regarding the factor structures of the AQ-50 and the AQ-26 is quite contradictory. As is evident from the discussion above, at least five distinct models have been championed by different researchers, with little consensus among them. This lack of consensus is understandable given the differences among the structural studies. For example, these studies used various factor analytic approaches: some used PCA, others EFA, and others switched to CFA. Further, some studies used dichotomous (0/1) scoring whereas others retained a four-point scale; however, no evidence was presented about the equivalence of these two scoring schemes. In addition, most studies (except for Hoekstra et al., 2008) did not account for the categorical nature of variables. Also, samples from different countries were used, further complicating the issue of generalizing the results to American student population. Moreover, only a single study (Hurst et al., 2007) investigated the AQ-26 as an independent scale, despite its potential

as a more parsimonious measure. This inconsistency among analyses makes it hard to draw a cohesive conclusion regarding the structure of the AQ-26. An additional study, therefore, is warranted.

The study conducted by Hoekstra et al. (2008) is the strongest of the structural studies on the AQ-50 or the AQ-26 because these researchers tested theoretically-specified models using the appropriate analysis – CFA with the estimation method accounting for the non-normal distribution of categorical variables. However, a three-factor model that has garnered both theoretical and empirical support was not tested by Hoekstra et al. (2008), leaving a gap in the literature. Also, a Dutch sample was used in Hoekstra et al.'s study (2008). Further, the AQ-26 was never considered in that study. Therefore, another set of CFAs needs to be conducted on an American sample in order to investigate all previously championed models, including the higher-order model championed by Hoekstra et al. (2008).

Importantly, researchers have used two different scoring schemes for the scale: four-point and two-point. Some studies used a two-point scoring scheme, whereas others used a four-point scoring scheme. However, the equivalence of the two scoring schemes has not been empirically investigated, despite the need for it. Establishing this equivalence is paramount for cumulating knowledge gained from the studies utilizing different scoring schemes.

In summary, the purpose of the current study was two-fold. First, the CFA analyses of previously championed models were used to contribute another piece of evidence regarding the dimensionality of the AQ-26. Second, results based on the two

scoring schemes were compared in order to provide an informed recommendation regarding the optimal scoring scheme.

In this study, the AQ-26 was administered on a computer to a representative sample of students at a southeastern public university. Notably, none of the previous studies used computer-based administration of the AQ. Upon collecting and screening the data, a series of CFAs were conducted using Mplus 5.2 in order to test model fit. Additional analyses followed as needed.

The subsequent chapters discuss the details of the present study in more detail. Chapter 2 provides an overview of the current understanding of Asperger's Syndrome and explicates the current state pertaining to measurement of autistic traits among adults. Also, Chapter 2 provides a more elaborative treatment to the structural studies conducted on the AQ. Chapter 3 outlines the methods employed in this research, including data collection procedures, participants, and the analysis. Chapter 4 presents the results of the analyses. Finally, Chapter 5 offers a discussion of the findings and suggestions for future research.

CHAPTER 2

Review of the Literature

Brief Overview of Asperger Syndrome

The recent upsurge in popular media coverage of autism and related disorders makes it hard to believe that autism has only been recognized in the clinical community since the 1940s (Attwood, 2007). Dr. Leo Kanner (1943) first described "autistic disturbances of affective contact" among children in a psychiatric facility. This initial account depicted severely affected patients with significant impairments in language, communication, socialization, and cognition. Not surprisingly, this seminal discovery provoked a stream of research targeted towards understanding and managing severe instances of autism. This extreme clinical picture painted by Kanner later acquired labels such as "classical autism" or "Kanner type autism" (Attwood, 2007). However, as the body of knowledge about autism expanded and early works were translated and disseminated, researchers began to recognize that autistic disorders exist on a continuum, ranging from severe, confounded with mental delay and other conditions, to the very mild, often perceived as eccentricity rather than abnormality.

One account in particular prompted researchers to consider autism as a spectrum disorder – that of Hans Asperger, a medical doctor in Vienna, Austria. Independently from Kanner, Asperger published his research based on 200 children in 1944 (Asperger, 1944). He also used the term "autism" and painted a similar clinical picture, although Kanner and he never met or exchanged correspondence. However, Asperger's clinical portrayal of "autistic psychopathy" was much milder than Kanner's, there was no cognitive or language delays. Nonetheless, children described by Asperger displayed key

autistic features, such as social impairments, difficulties with communication, and intense, narrowly-focused interests. Although the findings of these two clinicians surfaced around the same time, Asperger's work remained virtually unknown in the English-speaking world until the 1980s and 1990s. Lorna Wing was the first to summarize and publish Asperger's findings in 1981, followed by Uta Frith who translated Asperger's original paper (1991). Soon afterwards, the disorder gained recognition in the clinical community as Asperger's Syndrome (AS) or High Functioning Autism (HFA). It was formally included in the DSM-IV and ICD-10 in the 1990s (American Psychiatric Association, 1994; World Health Organization, 1990). Not surprisingly, scientific understanding of these developmental conditions is not as comprehensive as that of the other autistic disorders.

DSM-IV Diagnostic Criteria for Asperger Syndrome

AS is a lifelong pervasive developmental disorder that occupies a more able end of the autism spectrum (Attwood, 2007). It is not characterized by cognitive or language delays, but is typified by the autistic triad of impairments: social skills deficiencies, repetitive behaviors and communication difficulties. Consider the official diagnostic criteria stipulated by the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition-Text Revision (DSM-IV-TR; American Psychiatric Association, 2000). These criteria cover the core AS features: social interaction impairments and behavioral stereotypy. The following sections briefly describe the DSM-IV diagnostic criteria. The complete criteria can be found in Appendix B.

Social interaction impairment. The first criterion, qualitative impairment in social interaction, warrants special attention. Coined as a "primarily social disorder" (Schutte, 2008, p. 9), AS manifests itself mainly in the social realm. The DSM-IV instructs clinicians to look for impairments in at least two of the following: nonverbal behaviors (e.g., eye-to-eye gaze, facial expression, body posture), developmentally appropriate peer relationships, and spontaneous sharing of enjoyment, interests, or achievements.

Our world is laden with highly nuanced unwritten social rules. These rules come naturally to most, understood in the intuitive, rather than intellectual manner. Unable to abide by these disguised social rules, individuals with AS struggle with all aspects of social interactions. These aspects include, but are not limited to: social prompts, facial expressions, proximity to others, and gestures (Myles & Simpson, 2002). Many AS patients are able to compensate for this deficit by consciously deciphering the rules of social interactions and following prepared scripts. However, socializing never becomes natural to them and people with AS tend to commit social faux pas more often than others (Attwood, 2007). Being naturally inclined towards a literal interpretation of the world around them, individuals with AS often miss sarcasm, irony, and humor (Attwood, 2007).

One theory stipulates that AS challenges in the social arena stem from the lack of the "theory of mind", which is an understanding that others have varying thoughts and feelings and that one's actions affect others' thoughts and feelings (Attwood, 2007). Indeed, this inability to understand others can negatively impact reciprocal communication, such as turn-taking. Moreover, it can explain AS tendency to miss telltale signs indicative of people's emotions (i.e. voice intonations, frowns, smiles, body

language) and, consequently, fail to respond appropriately to such signs. For one reason or another, individuals with AS have impaired social skills that negatively affect many areas of functioning. However, a sole impairment in the social domain is not sufficient for the AS diagnosis.

Behavioral stereotypy. The next area of impairments pertains to the repetitive and stereotyped patterns of behavior and interests typical to individuals with AS. The latter can manifest in a strict adherence to routines or an intense interest in collecting or ordering objects (Attwood, 2007). Such interests usually fall in the area of folk physics (concerned with how things work) rather than in the area of folk psychology (concerned with how people work). They can represent a wide array of topics, such as mechanics, numbers, fabrics, geology, astronomy, mathematics, computer science, etc. (Myles & Simpson, 2002). It is not the area of interest that marks abnormality, but rather the intensity and stereotypy with which a person pursues a particular interest. Furthermore, individuals with AS often have exceptional rote memory capabilities and enjoy memorizing various trivia (Foster, 2002). Given that rote memory is typically a strong suit of people with AS and stress reduction appears to come from systematizing objects and information, it is possible that these intense interests bring enjoyment and satisfaction that social interaction fails to deliver. Not surprisingly, these intense hobbies, if successfully pursued, can potentially lead to rewarding professional occupations.

Other Features of Asperger Syndrome

The rest of the DSM-IV criteria call for eliminating the possibility of other disorders (e.g., schizophrenia, autism and pervasive developmental disorder) and specify

that these AS-typifying characteristics must actually hinder a person's functioning in order to be considered abnormal and not merely eccentric. Given that AS was only formally included in the DSM-IV and ICD-10 in the 1990s (American Psychiatric Association, 1994; World Health Organization, 1993), it is quite likely that the DSM-IV diagnosis as it currently stands is incomplete. In their discussion of the DSM-IV entry on AS, Myles, Cook, Miller, Rinner, and Robbins (2000) said that "the depth of knowledge is still so superficial that it will be years before we gain a full understanding of … AS characteristics" (p. 20). Only in the last decade, a number of breakthroughs have been made in understanding AS. These recent developments have allowed for identification of certain behavioral symptoms that appear to be typical to AS (Gillberg & Coleman, 2000). As Foster (2003) noted, broadening the scope of impairments beyond those included in the DSM-IV is essential for identifying individuals who might have AS. Furthermore, awareness of the latest research findings increases our chances to cover the full breadth of the construct. In short, other typology is known to be characteristic of AS and thus warrants our attention. The subsequent sections outline additional characteristics rendered by researchers to be common to AS. Those are: communication, executive functioning, sensory issues, and motor issues.

Communication. Unlike classical autism, AS is not marked by a severe impairment in communication (APA, 2000). According to the DSM-IV criteria, individuals with the suspected AS diagnosis exhibit "no clinically significant general delay in language" (APA, 2000). So technically speaking, communication competency conceptualized as language fluency is not a defining diagnostic criterion for the AS.

However, impairments in social interactions imply difficulties individuals with AS experience with the functional use of language for the communication purposes. Furthermore, researchers agree that individuals with AS typically have subtle difficulties with communication and language. These difficulties lie with pragmatic (as opposed to syntactic or phonological) features of language and communication (Baron-Cohen, Wheelright, Robinson, & Woodbury-Smith, 2005). Language skills of AS adults appear to be advanced at first - often characterized by rich vocabulary (probably due to the considerable time spent engaging in solitary activities, such as reading). However, their use of language is often marked by a peculiar pedantic characteristic. Howlin (2000) mentions that "abnormalities in reciprocity and pragmatics" are common for individuals with AS (p. 121). For example, they might engage in long monologues (especially if the topic is their special interest) or remain quiet for a very long time. Flat in affect or unbefitting in the context, the way individuals with AS use language to communicate seems artificial, somehow lacking the elusive quality of "naturalness".

Executive functioning. Executive functioning refers to one's ability to plan and carry out projects (Foster, 2002). Schutte (2003) describes it as "problem and goal identification, organization and planning, insight and awareness, initiation and modulation, and dexterity, flexibility, and speed" (p. 14). In other words, executive functioning entails coordinating a complex network of resources, skills, and constraints in order to achieve a previously set goal. A series of experiments comparing AS and neurotypical adults on the battery of executive functioning tests concluded that AS individuals are deficient in the areas of planning, organization, action monitoring,

initiation, intentionality, and goal/sub-goal coordination (Attwood, 2007). Furthermore, individuals with AS might be very inflexible when it comes to following the rules set up by somebody else and might get very offensive in response to criticism (Foster, 2003). This inability to accept criticism might jeopardize academic and professional success of individuals with AS, even though they might exhibit unusual creativity or intellectual ability.

Sensory issues. Sensory input (tactile, vestibular, proprioception, visual, auditory, gustatory, olfactory) is processed by AS individuals differently than by neurotypical populations (Dunn, Myles, & Orr, 2002). Although sensory idiosyncrasies appear to be common to people with AS, the specifics differ on a case-by-case basis. For example, some might be unable to tolerate certain smells, textures, lights, or noises (Attwood, 2007). Others might exhibit unusually high tolerance for heat, cold, or pain.

Motor issues. Individuals with AS often have unusual body posture, gait, pace, balance, and coordination (Smith, 2000). There may be issues with fine motor skills (such as required for handwriting) or overall clumsiness and accident proneness (Foster, 2002).

Prevalence of Autism and Asperger Syndrome

According to the latest report released in 2007 by the Center for Disease Control and Prevention (CDP), 1 in 150 8-year old children in multiple areas of the USA has an autism spectrum disorder. Assuming that this rate is accurate and has remained constant over the last two decades, we can estimate that about 560,000 individuals in the USA are affected. This rate is considerably higher than the estimates from 1970, which ranged from .07 to 2.4 per 100. Only in the late 20th-21st century, diagnosed cases increased from

4 per 10,000 to 5-6 per 1,000. It is undetermined whether this increase is due to the increased awareness of autism and related disorders, or an actual increase in prevalence. Besides, differences in the diagnostic criteria over time skew this estimate. A recent review of studies evaluating the prevalence of AS specifically concluded that AS prevalence ranges between 1 in 33,000 and 1 in 1,200 children. Unfortunately, no data are available regarding the prevalence of AS traits in adults.

Lack of research on adults with AS is especially concerning given that AS is recognized to be a life-long disorder (Attwood, 2007; Tantam, 2000). Unfortunately, current understanding of the AS traits among adults is especially limited, partially due to the lack of large-scale research studies with adult population (Tantam, 2000). More specifically, cognitive functioning of university students with AS represents an under-researched area (Myles & Simpson, 2002). This is unsettling given that there is reason to suspect that young adults pursuing college education and exhibiting mild autistic behaviors face unique challenges in both academic and social domains of the college life (Glennon, 2001; Smith, 2007). The next section further explains the potential difficulties facing students with AS in college and highlights the need to better serve this population.

Students with Asperger Syndrome

Adaptation to a university environment can be quite challenging for new students, regardless of their disability status. Newly acquired independence, academic workload, extracurricular activities, and social demands of college life all contribute to the adjustment-related stress of students. Fortunately, various formal and informal support services usually available on campus provide the necessary help to those seeking it.

However, students with AS represent an under-served population that may be struggling with the very essential components of college life yet may not be receiving adequate support from their home institution. Although few empirical research studies have investigated AS in college, a few topical reviews have examined the current knowledge of AS through the lens of the university experience. What we know about university environment, coupled with the current understanding of AS, can inform both theory and practice, and lay the groundwork for empirical investigations with a confirmative focus. Let us now consider how various features of AS might affect college adjustment.

Adjustment to College: Social Aspects

Several researchers have noted the potential difficulties encountered by a student with an impaired social ability in the college setting. In her discussion of the stress of the university experience for students with AS, Glennon (2001) focused on the importance of social interaction in the college settings. Indeed, college time offers increased opportunities for advancing social skills and building life-long social support networks. For example, there are informal social gatherings organized and attended by small groups of students, crowded sports and cultural events frequented by students and other members of the university. Aside from leisure, classroom time is often laden with social activities, such as discussions and team assignments. Consider the various ways in which students with AS might struggle in the social arena of college life.

Recall that AS is a primarily social disorder. It is characterized by difficulties understanding unspoken social “rules”, picking up on social cues, maintaining appropriate proximity to others, eye contact, posture, gestures, and so forth (Myles &

Simpson, 2002). Intact cognitive abilities allow youths with AS to master rudimentary social skills, such as quick greetings in the hallway. However, as social situations become increasingly prolonged and laden with simultaneous presentations of facial expressions, voice intonations, postures, and gazes, maintaining active social presence becomes very challenging for those affected with AS. As Myles and Simpson (2002) mention, youths with AS, unlike those with classical autism, *desire* social interaction and quickly become aware of their inadequacy in this domain. In turn, other psychological problems may arise, such as anxiety, depression, behavioral aggression, hyperactivity, emotional withdrawal, and low self-esteem (Glennon, 2001). Making the situation even worse, insipid facial and emotional expressions, coupled with social and communication barriers, prohibit those with AS from expressing their turmoil. In the similar vein, members of the university community may not recognize the distress of these students and assume that they are simply somewhat withdrawn and eccentric, but quite content.

Adjustment to College: Academics

Impairments in the social and executive functioning domains are also likely to affect academic performance. As briefly mentioned above, many higher-level college courses involve a good amount of teamwork. Designed to simulate real-life circumstances, these team projects are meant to teach students how to productively function within a group (Glennon, 2001). However, for somebody struggling with the more elementary aspects of social interaction, such a highly nuanced and socially demanding context presents an insurmountable challenge. Recall that individuals with AS experience difficulties with executive functioning, which pertains to one's ability to

manage time and resources, set both long-term and short-term plans, and generally organize and carry out projects (Foster, 2002). With that in mind, one can imagine the stressors stemming from the pressure to keep up with the academic calendars, homework assignments, class schedules, and paper deadlines. Unaccustomed to this level of organization skills required, students with AS might be overwhelmed with the executive demands and fall behind academically.

Aside from executive and social functioning, intellectual capabilities required to succeed in college warrant our attention. Although DSM-IV-TR criteria clearly states that intellectual and language abilities need to be intact for the AS diagnosis, the specific cognitive profile of AS is unknown (Myles & Simpson, 2002). In their review of existent evidence regarding academic and cognitive abilities of people with AS, Myles and Simpson (2002) summarize known intellectual strengths and weakness of individuals with AS. Expectedly, weaknesses lie in the areas of social judgment, visual-motor coordination, social mores, interpersonal situations, common sense, and social conventions. In addition, comprehension of abstract materials (i.e. metaphors, idioms) and inferentially-based materials presents another area of concern. At the same time, individuals with AS show an aptitude for non-verbal concept formulation, perceptual organization, and spatial visualization. In addition, comprehension of factual material is another strong suit of AS (Myles & Simpson, 2002). Unfortunately, research investigating the effect of the AS cognitive features in relation to academic success in college is not available. Instead, we rely on theoretical speculations and anecdotal accounts. Such speculations, however, are not meaningless, but rather informative. For

example, one can hypothesize about how special interests of students with AS might play out in college.

Recall that individuals with AS tend to have a restricted range of interests, with a few special interests which are atypical not in what they are, but rather in the intensity with which they are pursued. In college, this inclination towards specific subjects can be both a gift and a curse. Stereotyped interests may lie in different areas, such as geology, astronomy, mechanics, linguistics, or mathematics. The complexity and sophistication of these interests vary with one's intellectual ability; it appears that any topic rich in terminology can potentially be a special interest. If this special interest aligns with the direction of one's academic pursuit, a student with AS thrives – doing what he/she loves best, succeeding academically, and moving towards a future occupation. However, it is also possible that an obsessive occupation with the special interest can backfire due to the following factors. First, it might not be something that one could pursue academically and translate into an occupation. Second, poor executive functioning ability might lead a person with AS to pursue their interest at the expense of other responsibilities, such as other coursework. Third, younger students do not usually have the freedom to pursue their academic interests because they have to satisfy general education requirements, which cover an array of various subjects. Incidentally, this early stage of an undergraduate career is also the time when adjustment-related stress may be at its highest. It follows that educators should consider providing adjustments to fit the special interests of students with AS. This last point brings us to the next question: what are the accommodations currently available to students with AS?

Support Services

Research indicates that, despite a growing number of postsecondary students with AS, support services offered to these students are grossly inadequate. For example, a recent study conducted by Smith (2007) employed survey methods to explore the types of accommodations, services, and programs offered to students with AS at the postsecondary level. The researcher distributed a survey to 102 institutions, randomly selected out of 1,706 members of the Association on Higher Education and Disability (AHEAD) organization. Overall response rate was 28.4%, with the majority of participating institutions being universities. The majority of respondents (72%) reported that students with diagnosed AS were enrolled with the disabilities office. Services most commonly offered to students with AS included alternate testing sites, extra exam time, and individual advising. Other types of accommodations reported were reduced course load, environmental alterations (lighting and noise), and communication support, among others. Notably, a number of institutions reported that individualized career and personal counseling is available, albeit offered by counselors with little knowledge of AS. Expectedly so, little time and effort were reportedly devoted to educating faculty members about the syndrome. In summary, this study demonstrated that post-secondary students diagnosed with AS receive the same services as students with all other disabilities. This conclusion is unsettling, given that AS manifests itself quite differently from learning and physical disabilities. As Smith (2007) points out, “with a unique syndrome comes the need for unique accommodations” (p. 526).

In fact, such unique accommodations – specialized programs offered to college students with AS – do exist, although they are not widely available and lack empirical support. More and more specialized programs and guides designed for students with AS are now emerging, reflecting the need for such services. Concerns pertaining to such programs are multi-fold. First, these programs are only available at a considerable cost, thereby prohibiting students in low SES brackets from accessing such services (e.g., Achieving in Higher Education with Autism/Developmental Disabilities). Second, they are only available at select locations. The third and the most pertinent downside is that little evaluative research exists supporting the effectiveness of such programs (Smith, 2007). Such evaluative inquiry could improve the program and bolster support for its effectiveness. If, as a result of an independent evaluation, a specialized program is rendered to be a "reasonable accommodation" for this population group, then it should be freely available to students in need. In fact, reasonable accommodation for disabled students is a legally protected right, as stated in the US Federal Law (Americans with Disabilities Act, 1990, 2008; Grossman, 2001). However, given the grim state of research on AS in college, discussing legal issues might be premature.

As the last bridge to an adult life, college marks a distinct period in one's development. Higher education is not limited to retention of factual knowledge, but also strives to provide students with equal opportunities for "developing skills for adulthood, forming life-long relationships, identifying a vocational pathway, and participating in extracurricular activities" (Glennon, 2001, p. 185). If a disability places students at a disadvantage in attaining these goals, these students should receive adequate support

services. Students with AS have a set of very unique characteristics necessitating special accommodations. First, the impairments in the social domain hinder these students' success and well-being. Second, impaired executive functioning is likely to negatively affect one's academic performance. Finally, AS's special interests are likely to play a role in a student's life, although the nature of these effects is not well-understood. In sum, there is good reason to suspect that students with autistic traits but typical intelligence inconspicuously struggle in college. Further research, preferably on a large scale, should supplement our understanding of unique issues experienced by students with AS. This research could be used to inform specialized intervention and support programs targeted specifically for students with AS. AS might be a hidden disability, but its disguised nature is not an excuse for ignoring it. With the right support, these individuals have the potential to not only merely adjust, but also flourish in the academic environment. The question is: how can we identify those students with autistic traits?

Diagnosis of Asperger Syndrome

A diagnosis of a developmental condition is a rather complex endeavor requiring a synergy of various methods, such as standardized assessments, structured interviews, and clinical expertise. Of course, no single test in isolation can substitute a thorough clinical examination. In fact, the gold standard in clinical practice is expert clinical judgment, cultivated by a clinician through practice and research. However, quantitative screening tests are useful to clinicians because they provide objective evidence about the patient's symptoms. Screening instruments are even more useful to researchers whose goal is to explain the phenomenon as it occurs in the population as opposed to

understanding the individual patient. Such instruments need to be targeted towards identifying AS specifically and not autism in general. Although there are many different checklists, scales and interviews designed to aid in diagnosing autism in general, there are not nearly as many assessment instruments available for identifying AS specifically (Howlin, 2000). Reviewing all existing autism instruments is beyond the scope of this paper and is provided elsewhere (see Matson & Neal, 2009). Instead, I will briefly review assessment tools available for identifying AS, with a special emphasis on the adult population.

Cautionary Note

Before reviewing the screening instruments currently available for assessing AS in adults, a few words of caution are needed. AS was officially recognized as a clinical condition both in the DSM-IV and ICD-10 in the 1990s (American Psychiatric Association, 1994; World Health Organization, 1991). Since that time, a strong line of research dedicated to understanding this condition has emerged. This research movement most likely constitutes a healthy sign of scientific progression. However, this burgeoning research effort resulted in as much debate and controversy as it did in consensus. The debate mainly revolves around competing viewpoints regarding the right diagnostic approach (dimensional versus categorical), differences between Highly Functioning Autism and AS, and diagnostic criteria for AS (Howlin, 2000). The latter issue of the diagnostic criteria of AS halters the development of sound assessment instruments. In fact, in her review of assessment instruments for AS, Howlin (2000) argues that “in the absence of clear and clinically satisfactory diagnostic criteria, efforts to develop valid

assessment instruments may be attempting to put the horse before the cart" (p. 120). Since the publication of Howlin's review in 2000, researchers did not appear to agree on the one and only diagnostic criteria for the AS. However, a large number of various assessment tools emerged in the last decade, indicating that a collaborative effort to refine the diagnostic criteria has been launched. In the most recent review of the data based methods for assessing AS, Matson and Boisjoli (2008) recognize the emergence of such instruments as a good harbinger – signifying that "research teams are 'in the hunt' to develop reliable and valid diagnostic instruments" (p. 245). Indeed, several assessment instruments for assessing AS have since been developed. However, majority of these instruments are not suitable for screening for AS traits among college students.

Assessment Instruments for AS

Measures suitable for screening AS traits in a non-clinical population are few and warrant further study. Although there is a multitude of measures used to diagnose autism in clinical practice, none of these tools are appropriate for screening for AS traits among college students. Such a screening instrument needs to be confidential, efficient and non-intrusive to allow for large-scale collection of self-report data. Unfortunately, most of the clinical tools cannot be used for this purpose because of their format (interview protocols or informant-based measures) or targeted populations (children). In a recent review of assessment instruments for AS, Matson and Boisjoli (2008) provided a summary of such instruments. Out of 16 measures reviewed, only 5 were designed to target AS specifically and not autism. Out of these five, only three are suitable to use with adults (those over

18). To illustrate, let us consider a few diagnostic scales commonly used in clinical practice.

As previously mentioned, informant-based scales are commonly used in clinical practice. For example, the Gilliam Asperger Disorder Scale (GADS; Gilliam, 2001) is a standardized, norm-referenced, informant-based instrument designed to identify AS for individuals of ages 3 through 22. The scale consists of four subscales: (1) social interaction, (2) restricted patterns of behavior, (3) cognitive patterns, and (4) pragmatic skills. There is considerable evidence of inter-rater and internal reliability of GADS, as well as criterion-prediction validity evidence (Gilliam, 2001). However, as an informant-based test, this measure requires somebody close to a person with suspected AS to answer questions and therefore cannot be used in college settings without involving third parties. Moreover, the age range purportedly covered by the test (3 through 22) makes one suspect whether the instrument functions equally well for individuals of all these ages.

Another common way to assess AS in clinical practice is to use interview protocols. For example, the Asperger Syndrome Diagnostic Interview (ASDI; Gillberg, Gillberg, Rastam, & Wentz, 2001) is a structured clinical interview which also requires cooperation of an informant. The ASDI was designed to diagnose adolescents and young adults with suspected high-functioning autism spectrum disorder, including, but not limited to, AS. Preliminary clinical data provide adequate evidence regarding inter-rater reliability, test-retest stability, and diagnostic validity (Gillberg et al., 2001).

Unfortunately, the ASDI is not a self-report measure and requires raters with clinical expertise.

Need for a Self-Report Instrument

Diagnostic tools exemplified above are not a good match for understanding the distribution of AS traits in the general student population because they tend to require one-on-one administration by a trained clinician or an interview with an informant. Therefore, they cannot be used for research purposes in college settings. To serve this purpose an instrument is needed that will allow information to be collected from a large number of students. Such a self-report measure needs to be brief, confidential, non-intrusive, and psychometrically sound.

The benefits of such an instrument include, but are not limited to: (1) enabling researchers to study AS traits in relation to related constructs and co-morbid disorders, such as social anxiety and depression; (2) gauging the prevalence of AS traits in the general population; (3) estimating the prevalence of AS traits among different groups (e.g. males and females, scientists and non-scientists); and (4) screening for AS in clinical settings. Several researchers have developed self-report measures of AS to meet this need.

Self-Report Measures of Asperger Syndrome

There are currently three published self-report measures purportedly assessing mild autistic traits among adults of typical intelligence. The sections below review these instruments and evaluate the need to study them further. Scale development process undertaken by the authors of these instruments is briefly described. Further, the suitability

of these instruments for the use in research and practice is evaluated based on the existing literature.

Ritvo Autism and Asperger's Diagnostic Scale (RAADS)

Ritvo Autism and Asperger's Diagnostic Scale (RAADS; Ritvo, Ritvo, Guthrie, Yuwiler, Ritvo, & Weisbender, 2008) is a 78-item self-report scale designed as a diagnostic tool for identifying AS and HFA in adults, based on the DSM-IV-TR and ICD-10 criteria. The authors wrote items tapping into three domains: 1) social relatedness, 2) language and communication, 3) sensorimotor and stereotypies. Notably, the Likert scale used for this questionnaire is developmental in nature, with the following categories: "true now and when I was young", "true only now, true only when I was young", and "never true". Sixty items purportedly tap into autistic features (e.g., "It is very difficult for me to understand some emotions"), with 18 items being negatively worded (e.g., "I can tell when someone says one thing but means something else"). The authors administered RAADS to 8 individuals with autism, 8 with AS, and 16 controls. Group comparisons indicated that individuals with AS and autism scored significantly higher than controls on all RAADS subscales, and on all but one items. Reliabilities for the hypothesized subscales (Cronbach's alphas) ranged from 0.60 to 0.86. However, the scale was not factor analyzed so the empirical evidence supporting the factor structure of RAADS scores is still lacking. Although RAADS is a promising addition to the field, empirical understanding of this scale's functionality is very scarce as only a single, original study exists (Ritvo et al., 2008).

Adult Asperger Syndrome Scale (AASS)

Another self-report questionnaire had been developed, although it is rarely mentioned in the autism literature. Adult Asperger Syndrome Scale (AASS; Foster, 2003) is a 95-item scale designed as a diagnostic instrument for adults with AS. It can be filled out either by a person with suspected AS or somebody close to him/her. Exploratory factor analysis ($N = 196$) revealed that six factors are likely to underlie the AASS scores. These factors are: (1) communication, (2) executive skills, (3) social skills, (4) memory, (5) sensory issues, and (6) self-absorption. These results are tentative due to the use of orthogonal rotation, which means that factors were not allowed to correlate with each other. Because other authors stipulate that AS traits are likely to be related to each other (Austin, 2005), the use of orthogonal rotation by Foster is questionable. Foster (2002) developed this instrument in hopes that sufficient reliability and validity evidence would be collected in the future to qualify this scale as a diagnostic instrument. Unfortunately, no other studies using the AASS have been disseminated. Consequently, evidence as to the functionality of the AASS is very limited, prohibiting the use of the AASS in research and practice.

Autism-Spectrum Quotient (AQ)

Finally, the Autism-Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) is a 50-item self-report questionnaire designed specifically to assess autistic traits in a non-clinical adult population. Unlike the other instruments reviewed, the AQ is a screening tool, rather than a diagnostic tool. The authors warn against the use of any single instrument for the clinical diagnosis. In fact, they later developed a diagnostic system, “The Adult Asperger’s Assessment (AAA): A

Diagnostic Method", consisting of a questionnaire covering the DSM-IV criteria, the AQ, and the Empathy Quotient Scale (Baron-Cohen, Wheelright, Robinson, & Woodbury-Smith, 2005). Before AAA can be used as a legitimate diagnostic system, supporting reliability and validity evidence of the instruments comprising this system needs to be acquired. The AQ has received considerable empirical support, compared to other scales in this family. In fact, a number of studies have investigated the psychometric properties of the AQ (e.g., Stewart & Austin, 2009). Functionality of the AQ has been examined using British (Baron-Cohen et al., 2001; Austin, 2005), Austrian (Voracek & Dressler, 2006), Japanese (Wakabayashi et al., 2006), and American samples (Hurst et al., 2007). Further, the AQ has been used in research to examine relationships of AS traits with the big five personality traits (e.g., Austin, 2005), schizotypal personality traits (Hurst, Nelson-Gray, Mitchell, & Kwapil, 2006), differences between scientists and non-scientists (Hoekstra, Bartels, Cath, & Boomsma, 2008) and sensitivity (Liss, Maillopux, & Erchull, 2008).

The original instrument consists of 50 items and is more commonly used in research (the AQ-50). The short version identified by Austin (2005) consists of 26 items (the AQ-26). Although investigations of the relationships between the two AQ scales and other variables can provide fruitful validity information, this line of research might be somewhat premature, given that the dimensionality of both scales is still unclear. Specifically, the AQ-26 lacks empirical support despite its potential as a more parsimonious measure. For this reason, the current study focused on the AQ-26, while taking previous research on both AQ-50 and AQ-26 into account.

Validity Evidence for AQ-50 and AQ-26

According to the *Standards for Educational and Psychological Testing* (1999), validity of an instrument refers to "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (p.9). Importantly, validity is not an all or nothing property of a test, but rather a body of empirical evidence supporting the interpretation of test *scores.* In particular, construct validation refers to the cyclical, multi-stage process of determining whether the instrument measures the construct it purports to measure. As such, it is imperative to gather construct validity evidence for an instrument before using it in research or practice.

The strong program of construct validation suggested by Benson (1998) provides a useful framework for scale development. This program consists of three stages: (a) substantive, (b) structural, and (c) external. The substantive stage entails defining both theoretical and empirical domains of the construct. Benson cautions the reader to two pitfalls commonly occurring in this stage: construct irrelevance (i.e. describing dimensions *unrelated* to the focal construct) and construct under-representation (i.e. omission of dimensions *directly related* to the focal construct). The structural stage refers to the internal domain investigations, such as dimensionality and reliability of the scale. Finally, the external stage involves studies focused on the nomological network of the construct, which is a theoretical map of its relationships with related variables. The external stage subsumes group differentiation and correlation studies. This final stage is the most informative and, arguably, the most important stage of construct validation as it provides the most compelling evidence that the instrument measures what it purports to

measure. However, no single stage of construct validation, regardless of how thoroughly investigated, suffices as evidence for test validity if considered in isolation from the other stages. As such, stages of construct validation comprise a validation *process* in which one stage informs another.

In accordance with Benson's (1998) strong program of construct validation, the following stages of construct validation will be examined in the next sections: (a) the substantive stage, which pertains to the theoretical and empirical domains underlying the scale, and (b) the structural stage, which pertains to the internal structure of the scale. The third stage in the Benson's framework involves evaluating the external validity of the scale, or the relationships with related constructs and the scale's capability to differentiate between groups. However, the external stage was not be examined in the current study because the structural stage needs substantial work and must be thoroughly addressed first. Collecting empirical evidence for the structural validity of the instrument *prior* to examining how its scores relate to other variables is very important. Psychometric studies can shed light on the dimensionality of the scale and suggest the correct scoring procedure (total score versus subscales). In addition, research focusing on the internal domain of the AQ can provide information about the reliability of the subscales and quality of the items. Before proceeding to discussing the psychometric studies of the scale, it is necessary to consider the substantive validity evidence and the scale development process.

Development of the AQ

The first of its kind, the AQ-50 was created by a team of researchers and practitioners at the Autism Research Center in London, Great Britain in 2001. Baron-Cohen and his colleagues recognized the need for a brief self-report measure to screen for autistic traits and developed a 50-item measure using current scientific understanding of AS as well as their own clinical expertise. The self-report was considered appropriate because individuals with AS traits do not suffer from a cognitive impairment and are thus fully capable of accurate responding. Initial scale development underwent rigorous procedures, including comprehension checks, piloting the instrument on adults with and without disabilities, and item revision and deletion. In general, the authors were guided by the theoretical assumption that autistic traits "lie on a continuum of social-communication disability, with AS as the bridge between autism and normality" (Baron-Cohen et al., 2001, p. 6). Given this conceptualization of AS, the authors felt that it is was reasonable to create a quantitative measure. Further, they worded some of the items in terms of preference rather than ability in case some individuals are incapable of accurately judging their own abilities (e.g., some of the item stems started with "I enjoy…" or "I find it easy to…" instead of "I am good at…"). Approximately half of the items were negatively worded. The response scale is a four-point Likert scale ranging from "definitely agree" to "definitely disagree".

Although the respondents use a 4-point Likert scale to respond to items, the test developers recommended scoring of categories 1 (definitely agree) and 2 (agree) as 1 and of categories 3 (disagree) and 4 (definitely disagree) as 0. The authors also recommend that a simple sum of the 50 dichotomously scored items be used as the AQ-

50 total score, with higher values purportedly indicating heavier autistic load. It should be noted that the original 4-point scale is retained in the scoring of the AQ-50 by some researchers (e.g. Stewart & Austin, 2005).

Unfortunately, the test developers did not empirically investigate the scale's dimensionality. Investigating the dimensionality underlying AQ responses is imperative to understanding whether the item responses relate to each other in a way that supports the authors' conceptualization of the scale. Moreover, it informs researchers on how to appropriately score the measure. For example, if five distinct dimensions are found to underlie the AQ responses, then five different scores should be calculated. If the scale is uni-dimensional, then a single score can be calculated. If the higher-order factor is found to subsume a number of lower-order factors, then both a total and sub-scale scores can be meaningfully interpreted. Several competing structural models have been hypothesized by different researchers to underlie the AQ. Some of these models are couched in theory, whereas others are empirically derived. The subsequent sections present these different models in the order in which they appeared in research chronologically.

Structural Stage of AQ-50 and AQ-26

Model 1: Single-Factor

The scale was created to assess a cohesive set of symptoms altogether comprising a developmental condition known as Asperger's Syndrome. Although the test developers conceptualized AS to be multi-dimensional and built the scale in accordance with such a conceptualization, they still recommended calculating a total score. However, the total score cannot be meaningfully interpreted unless a single latent variable drives the

responses to all items. In order to investigate the single-factor hypothesis, a uni-dimensional model consisting of a single latent variable *Autistic Load* was specified. This model is presented in Figure 1.

Model 2: Five-Factor (Baron-Cohen et al., 2001)

Baron-Cohen et al. (2001) crafted the items based on the autistic "triad" of impairments (APA, 2000). This triad consists of social skills deficiencies, repetitive behaviors and communication difficulties. A more detailed description of the triad is presented earlier in this work. Also, the DSM-IV-TR diagnostic criteria for AS are outlined in Appendix B. Notably, communication difficulties are listed in the DSM-IV-TR criteria as part of the social skills deficiencies and not as an independent criterion.

In addition to the diagnostic criteria outlined above, the authors also referred to the "other demonstrated areas of cognitive abnormality in autism" when drafting the items (Baron-Cohen et al., 2001, p. 6). Unfortunately, the authors did not provide further detail about which areas of impairment guided item creation. Instead, they proposed that the AQ addresses the following five inter-related categories, with 10 items assigned to each: (1) *Social Skill*, (2) *Attention Switching*, (3) *Attention to Detail*, (4) *Communication*, (5) *Imagination*. This model is graphically depicted in Figure 2. In addition, Table 1 provides an item-factor mapping of the five-factor model, along with the other models described later on in this work. Although it is clear that theoretical foundations guided the development of the AQ-50, it is less clear how exactly the proposed dimensions map onto the diagnostic criteria and other areas of impairment. Upon constructing the AQ-50, Baron-Cohen et al. administered the measure to adults diagnosed with an autistic disorder

(N = 58), college students without known impairments (N = 840), adult volunteers (N = 174), and winners of the Mathematics Olympiad (N = 16) and considered group differences on the proposed AQ-50 subscales. Although the pattern of scores across groups support the validity of the AQ scores, no factor analysis was conducted to provide evidence about the dimensionality of the AQ. Several researchers have recognized the need for conducting such analyses to provide further evidence on the structural validity of the AQ. These structural studies are outlined below.

Model 3: Three-Factor (Austin, 2005; Hurst et al., 2007)

Austin (2005). The first structural study of the AQ-50 was conducted by Austin (2005) using a sample of British undergraduate students (N = 201). This author conducted a principal component analysis (PCA) with oblique rotation (allowing components to correlate) on the AQ-50. Based on this analysis, Austin supported a three-component solution (explaining 28% of the variance), although the scree plot indicated that a one-component solution was also plausible (explaining 14% of the variance). Unfortunately, Baron-Cohen's five-component model was not explored in this analysis.

Only 26 items with pattern coefficients above 0.4[2] were reported (indicating that at least 16% of the variance in these items was explained by the corresponding component). Austin's (2005) article prompted others to consider these 26 items as a stand-alone scale, here referred to as the AQ-26. Notably, the model championed by

[2] Although Austin only reports those items with pattern coefficients above 0.4 and separates items by component in a table, it is unclear whether any of the items had split loadings.

Austin and described below was based on the analysis of the full AQ-50, and not AQ-26. This model consisted of three components: (1) *Social Skills,* (2) *Details/ Patterns,* (3) *Communication/ Mindreading*. Table 1 presents these three factors and their corresponding items. Internal consistency estimates[3] (coefficient alpha α) based on only those 26 items that had at least a 0.4 pattern coefficient on each component were, respectively, 0.85, 0.70, 0.66. The first and third components correlated moderately at 0.2; other inter-component correlations were not reported. This model gained some support in the next psychometric analysis of the AQ-26, which explored the 26 items separately. The three-component model is presented in Figure 3.

Hurst et al. (2007). The second structural stage study of the AQ-50 and the AQ-26 was conducted by Hurst et al. (2007) using an American college student sample ($N = 1005$). Similar to Austin's study, PCA with oblique rotation was employed to explore the components of both scales. Unlike Austin (2005), however, these researchers used dichotomous scoring (0/1). Baron-Cohen's five-dimensional model of the AQ-50 was explored and deemed implausible because the items did not "load" on the factors specified by the model; about 27% of the variance was explained by this model. Although the scree plot suggested both a one- and three-component solution, these solutions were not explored by the authors. In fact, the scree plots associated with the PCA of both the AQ-50 and AQ-26 (described next) were reported, but not used to guide the decision as to which component solution to explore.

[3] It is assumed that Austin calculated these coefficients using only these 26 items with pattern coefficients above 0.4, although it is unclear from the article.

Next, the AQ-26 (items reported in Austin's study) was subjected to the same analysis. The scree plot indicated the plausibility of a 1- or 3-component solution. Unfortunately, neither the single-component nor the five-component models suggested by the results were evaluated on the AQ-26. However, the three-component model previously endorsed by Austin (2005) was explored on the AQ-26, explaining about 29% of the variance. Hurst et al. (2007) concluded that their results "closely resemble those of Austin (2005) and support a three-factor solution" (p. 1946). However, several items (40, 34, 50, 43, 25, 39, 7, 37) had item loadings[4] of less than 0.4 and another item cross-loaded on the first two components. Nonetheless, Hurst et al. (2007) supported a three-component solution first championed by Austin (2005) and used the same labels for the three components (see Figure 3). The internal consistency estimates (α) for these subscales were, respectively, 0.75, 0.54, 0.42. Curiously, Hurst et al. reported the reliability index (α) for the total AQ-26 (0.63), although a uni-dimensional model was not considered in this study. Overall, Hurst et al. (2007) study garnered some empirical support for the three-dimensional model postulated by Austin (2005). Furthermore, the three-dimensional model also has some theoretical support because it aligns with the autistic triad of impairments comprised of (1) social impairments, (2) repetitive behaviors, and (3) impaired communication (APA, 2000).

Model 4: Higher-Order Factor (Hoekstra et al., 2008)

[4] All loadings were reported, but it is unclear whether these loadings are pattern or structure coefficients.

In the search for more evidence on the psychometric properties of the AQ-50, a team of Dutch scientists (Hoekstra, Bartels, Cath, & Boomsma, 2008) translated the AQ-50 into Dutch and administered it to a sample of Dutch college-aged students without known disabilities (N = 961) and a general population sample (N = 302). They conducted confirmatory factor analysis (CFA) with diagonally weighted least square (DWLS) estimation to assess the fit of the competing factor structure models hypothesized to underlie the AQ-50 separately in each sample. These authors used the 4-point scoring scheme for the items. A few important conclusions regarding the factor structure of the AQ-50 were made in this study. First, the uni-dimensional model did not yield adequate fit in either one of the samples. Despite this finding, the reliability estimate for the total AQ score was reported (α = 0.81 in the student sample) and the total score was used in the subsequent analyses. Secondly, the authors found that the five-dimensional model proposed by Baron-Cohen et al. (2001) was not an appropriate fit for the data either. Third, the authors proposed and championed an alternative model, which consisted of a higher-order factor (1) *Social Interaction* (α = 0.84) subsuming four lower order factors (*Social Skill* (α = 0.76), *Attention Switching* (α = 0.63), *Communication* (α = 0.52) and *Imagination* (α =0.63) and one separate factor (2) *Attention to Detail* (α = 0.68). The higher-order and the separate factor correlated at 0.19. The higher-order model of AQ-50 supported in this study was also deemed plausible for the AQ-26, although no CFAs were conducted on the shorter scale in this study. The graphical depiction of this higher-order model specified based on Hoekstra et al. (2008) research is presented in Figure 4.

Model 5: Four-Factor (Stewart & Austin, 2009)

Still lacking conclusive evidence regarding the factor structure of either version of the AQ, Stewart and Austin (2009) conducted another set of psychometric analyses using a British college-age student sample ($N = 536$) and the full length version of the instrument (AQ-50). These researchers carried out an EFA analysis within a CFA framework, which allowed them to get fit indices for multiple exploratory models. In this study, the 4-point scoring scheme was used. First, they conducted an exploratory factor analysis (EFA, with oblique rotation) and based on a scree plot concluded that four, six, and eight factor solutions were all plausible. The four-factor solution explained 29% of the variance. Next, a series of CFA analyses were carried out on the same sample to test the fit of 1 to 8-factor solutions[5]. Based on the CFA results, the authors championed a four-factor model consisting of the following four factors: (1) *Socialness,* (2) *Patterns,* (3) *Understanding Others/ Communication,* (4) *Imagination.* Reliability indices were reported for the *Patterns* factor ($\alpha = 0.69$) and the *Imagination* factor ($\alpha = 0.55$), and the other alphas were said to exceed 0.70. Notably, *Socialness* and *Patterns* correlated at 0.16; *Socialness* and *Understanding others/Communication* correlated at 0.22; and *Imagination* and *Understanding others/ Communication* correlated at 0.18. Notably, the inter-factor correlations, as well as reported pattern coefficients, were those resulting from the EFA study. A four-factor inter-correlated model of the AQ-26 was specified based on Stewart and Austin's (2009) model. This model is presented in Figure 5.

Inconsistencies Among Models

[5] It is unclear whether all items were allowed to load on all factors in these analyses or if the authors restricted the loadings in some way based on the EFA study.

The models outlined share some similarities but also differ in notable ways. Table 1 includes item-factor maps as specified by different models. The five-factor model originally hypothesized by Baron-Cohen et al. (2001) (Figure 2) is nested within a single-factor model (Figure 1). The higher-order model championed by Hoekstra et al. (2008) (Figure 4*)* merges four of the factors outlined in the five-factor model under a single higher-order (*Social Skills)* and delineates one factor as separate (*Attention to Detail).* As such, item-factor mapping in the five-factor and higher-order factor models is the same; only the higher-order factor model is depicted in Table 1. In contrast, the three-factor model supported by Austin (2005) and Hurst et al. (2007) (Figure 3) is not a simple merging of factors from the five-factor model and is therefore not nested within the five-factor model. That is, items from the five original factors load differentially on the three factors. Notably, all models, except for the single-factor model, include three dimensions, with some of the items consistently mapping onto them.

The four-factor model championed by Stewart and Austin (2009) (Figure 5) is very similar to the three-factor model, with the exception of items 40 and 50 loading on the fourth factor *Imagination. Imagination* factor does not seem to be very strong as there are only two items mapped to it (40 and 50). These two items both talk about playing with children, making the reader suspect whether these two items really tap into imagination, or just one's inclination to playing with children.

With the exception of the *Imagination* factor, three-factor and four-factor models are identical. However, some items appear to "travel" across dimensions when compared

to the five-factor solution. Let us consider item mapping of these three factors and their correspondence with the original five-dimensional model (refer to Table 1).

Communication factor appears to be the most problematic. Out of the six items originally assigned to *Communication* in the five-factor model, only three items (7, 35, 39) load on the same factor in the other models. The other three items (17, 26, 38) tend to load on the *Social Skills* factor.

Attention Switching factor might not be distinct from *Attention to Detail* factor because five of the seven items originally assigned to *Attention Switching* consistently load on *Patterns/Details* both in three-factor and four-factor models, with the *Patterns/Details* factor also consisting of *Attention to Details* factor.

Social Skills factor appears to be the most stable across solutions with six out of the original seven items consistently mapping onto the *Social Skills* factor. One item (45) might be problematic as it tends to load on the *Communication* factor in the three- and four-factor solutions.

Limitations of the Previous Studies

Although the psychometric analyses conducted on AQ-50 and AQ-26 thus far are very informative, a few caveats need to be noted. These caveats have important implications for future research studies, of both structural and substantive nature. The next sections outline these problematic areas.

Two Scoring Schemes

An important inconsistency across the structural studies lies in two different scoring schemes used. This inconsistency across studies further exacerbates the issue of

generalizing the results. Let us now consider the two different scoring schemes used to score the instrument.

The first one can be referred to as a 4-point scheme. Recall that the AQ-26 has a 4-point Likert-type scale ranging from "1" ("Definitely Disagree") to "4" ("Definitely Agree"). Some researchers retain this 4-point scale when conducting psychometric analysis of the AQ. For example, Austin (2005), Hoekstra et al. (2008), and Stewart and Austin (2009) all retained the 4-point scale. Notably, the 4-point scoring scheme results in a categorical variable, with too few categories to be considered continuous (Finney & DiStefano, 2006). Others, however, use a different scoring scheme when conducting structural studies of the AQ.

The second scoring scheme can be labeled as a 2-point or dichotomous scheme. Baron-Cohen et al. (2001) recommended scoring categories 1 (definitely agree) and 2 (agree) as 1 and categories 3 (disagree) and 4 (definitely disagree) as 0. This recommendation was made to facilitate the ease of scoring. This scoring scheme results in a dichotomous variable. Some researchers, including Hurst et al. (2007) adhered to this recommendation.

Having these two different scoring schemes presents a problem. First, the rationale originally provided for the 2-point scoring scheme was the ease of scoring. However, it is unclear why the scale is administered on a 4-point scale in the first place if the categories are collapsed for the analysis. If provided with only two categories, participants might respond differently to the items. Also, dichotomizing the variables likely leads to a loss of information. Although Baron-Cohen et al. (2001) stated that

analyzing the data retaining the four-point scale yielded the same pattern of results overall, they did not present any evidence corroborating this statement. Unfortunately, the other researchers using the scale did not compare these two scoring schemes. Consequently, both schemes are being used in research but it is unclear whether the results of psychometric studies using the 4-point scheme are applicable to the use with a 2-point scheme. So the questions still remain: do these two scoring schemes yield the same results and are therefore interchangeable? If not, what are the implications of using one scoring scheme as opposed to another?

EFA versus PCA versus CFA

Various factor analytic techniques have different purposes. Although they can inform each other, they cannot be compared side by side. Unfortunately, various different factor analytic methods were used to analyze the AQ. For example, exploratory techniques were used in the first few studies (PCA or EFA). Such analyses are, as the name implies, exploratory – they can shed light on the plausible underlying dimensions of the scale, but cannot be used to test the fit of models specified a priori. Although both analyses can be used to gather evidence for the substantive stage of construct validation, CFA techniques are more appropriate when specific models are postulated because CFA provides a more rigorous test of the fit of the models to the data. Also, PCA is not equivalent to EFA because the former does not account for measurement error, leading to biased parameter estimates (Benson & Nasser, 1998; Preacher & MacCallum, 2003). It should be noted that PCA and EFA results do tend to converge when a large number of items are analyzed. Therefore, it might be the case that PCA and EFA of AQ-50 yield

similar results. Despite the equivalence between PCA and EFA likely in this situation, the two analyses have two different worldviews that have been largely ignored by researchers using PCA. PCA yields components, whereas EFA yields factors. Unlike factors, components cannot be interpreted as latent underlying constructs driving the responses to the items. PCA is best fitted for data reduction and yields components which are just linear combinations of observed variables, not latent factors (Preacher & MacCallum, 2003). It is important to interpret the results of these analyses accurately in order to avoid misleading the reader.

It appears that several studies interpreted the results of a PCA as if they were EFA results (Austin, 2005; Hurst et al., 2007). Throughout their articles, these authors refer to components as factors. It is unclear which exploratory method was used in the most recent study (Stewart & Austin, 2009). In general, the choice of the statistical technique should align with the theoretical question. It might be that PCAs were appropriately used in some analyses, but theoretical justifications for these choices were not provided.

Methods for Categorical Data

Notably, both 4-point and 2-point Likert scale scoring results in categorical variables, warranting appropriate factor analytic techniques that account for non-normally distributed data and non-linear relationships between item responses and factors (Finney & DiStefano, 2006). However, only a single study accounted for the categorical nature of the 4-point Likert scale responses by using the appropriate estimation method (DWLS) (Hoekstra et al., 2009).

Cross-National Samples

Although the AQ-50 and AQ-26 are being used in different countries, the cross-national equivalence of the measure has not yet been established. Measurement invariance studies are commonly conducted to establish cross-national equivalence (Steenkamp & Baumgartner, 1998). Thus, it is most informative to conduct further analyses of a measure after functionality of this measure is found to be comparable across different populations. Unfortunately, measurement invariance of the AQ-50 or the AQ-26 has not been examined. This concern is especially pertinent when translated versions of the instrument are used (as is the case with Dutch version of the AQ-50).

A Priori versus Post Hoc Model Specification

CFAs are best used for testing theoretical models firmly couched in previous research and specified a priori (Kline, 2005). If used for testing models empirically specified post hoc, CFA results are not as compelling. This is because post hoc model specifications capitalize on chance and idiosyncrasies of a given sample, and yield results that are unlikely to replicate in an independent sample (Boomsma, 2000; Kline, 2005).

Unfortunately, it is not always clear whether the researchers specified the models a priori or post hoc. For example, the study conducted by Hoekstra et al. (2008) might have issues associated with post hoc model specification. Unfortunately, it is not immediately evident whether the higher-order model championed by Hoekstra et al. (2008) was specified a priori or post hoc. Further, the theoretical conceptualizations of AQ-50 dimensionality described in the introduction of Hoekstra et al. (2008) article were never translated into statistical models and the fit estimations for these models were never reported. Sadly, Hoekstra et al. (2008) did not test the three-factor model that has

garnered some theoretical and empirical support and thus could be specified a priori (Austin, 2005; Hurst et al., 2007).

Fit Indices

Recent simulation studies exploring the functionality of different fit indices revealed that some indices perform better than others do when it comes to identifying simple and complex model misspecifications. Some fit indices were not recommended because they were not sensitive to model misspecification and sensitive to sample size (Hu & Bentler, 1998; 1999). Also, cutoff values for identifying model fit were provided. Unfortunately, researchers who conducted CFA on AQ-50 did not follow these recommendations and interpreted some fit indices that are likely to produce biased values, and overlooked the cutoff recommendations. The following section describes recommendations concerning fit in more detail.

The following fit indices reported by Hoekstra et al. (2008) are not recommended. Both GFI and PGFI tend to be not sensitive to model misspecification and sensitive to sample size, and are thus not recommended as they might lead to biased conclusions (Hu & Bentler, 1998). Similarly, the ECVI (which estimates the fit of a model in a comparable sample) is not recommended for the same reasons (Hu & Bentler, 1998). Also, the fit index recommended for the use with the categorical data (WRMR) and a summary of standardized residuals (which is informative for diagnosing model misfit) was missing.

Moreover, the evaluation of fit indices reported by Hoekstra et al. (2008) do not lead to the conclusion that either one of the models were plausible, according to the

recommended cutoff values. For example, the SRMR values did not approach a 0.08 cutoff and GFI did not approach a 0.90 cutoff (as recommended by Hu & Bentler, 1998; 1999).

Internal Consistency

Numerous studies report the internal consistency estimates for the AQ subscales (for different models), which range from poor to adequate to good. However, interpreting reliability estimates is premature before the factor structure of the instrument is established. As well, such estimates are overly conservative if reported for the factors that are not congeneric.

Purpose of the Current Study

Overall, the evidence regarding the factor structure of AQ-50 and AQ-26 is quite contradictory. As evident from the discussion above, at least five distinct models have been championed by different researchers, with little consensus among them. This lack of consensus is understandable given the differences among the structural studies. For example, these studies used various factor analytic approaches: some used PCA, others EFA, and others switched to the CFA. Further, some studies used a 2-point scoring whereas others retained a 4-point scale; however, no evidence was presented about the equivalence of these two scoring schemes. In addition, most studies (except for Hoekstra et al., 2008) did not account for the categorical nature of variables. Also, samples from different countries were used, further complicating the issue of generalizing the results to American student population. Moreover, only a single study (Hurst et al., 2007) investigated the AQ-26 as an independent scale, despite its potential as a more

parsimonious measure. This inconsistency of analyses makes it hard to draw a cohesive conclusion regarding the structure of the AQ-26 and warrants another study.

The study conducted by Hoekstra et al. (2008) is the strongest out of the structural studies on AQ-50 or AQ-26 because these researchers tested theoretically specified models using the appropriate analysis – CFA with the estimation method accounting for the non-normal distribution of categorical variables. However, a three-factor model that has garnered both theoretical and empirical support was not tested by Hoekstra et al. (2008), leaving a gap in the literature. Also, a Dutch sample was used in Hoekstra et al.'s study (2008). Further, the AQ-26 was never considered in that study. Therefore, another set of CFA needs to be conducted on an American sample in order to investigate all previously championed models, including the higher-order model championed by Hoekstra et al. (2008).

Importantly, researchers tend to use two different scoring schemes for the scale: 4-point and 2-point. The studies using two different scoring schemes cannot be compared side by side. The equivalence of the two scoring schemes has not been empirically investigated, despite the need for it.

In summary, the purpose of the current study was two-fold. First, the CFA analysis of previously championed models served as another piece of evidence regarding the dimensionality of the AQ-26. Second, the results based on the two scoring schemes were compared in order to provide an informed recommendation regarding the optimal scoring scheme.

CHAPTER 3

Methods

Procedure

Data were collected during the fall semester of 2009. Responses were gathered through computer-based administration (Qualtrics survey software) of a battery of psychological measures. Participants had to select a response for every one of the items in order to proceed (i.e., the forced-choice response option of the survey software was enabled). This feature alerted those who accidentally missed an item to provide a response and minimized missing data. All participants provided informed consent to participate. Specifically, all students were informed that participation in the study was voluntary, that participant confidentiality would be protected at all costs, and that the aggregate results might be disseminated through publications or presentations. At the beginning of each session, participants were informed that the purpose of the study was to develop better measures of student affective and cognitive development. However, students were not informed that the purpose of the study was to investigate autistic traits among students. This was done in order to ensure that attitudes toward autism or disability status would not affect students' responses. All participants were encouraged to contact the researcher for full debriefing in spring 2010.

The questionnaires were administered in a controlled environment to a limited number of students at a time (maximum = 50) in one of the computer laboratories on campus. Overall, data were collected in several testing sessions. The primary researcher, who served as the proctor, monitored every testing session and followed a standardized protocol to ensure that participants devoted adequate time and effort to the

questionnaires. The proctor was present at all times during the data collection to field questions, solve technical problems, and ensure that participants exerted motivation while taking the tests.

Participants

The sample consisted of undergraduate students enrolled in psychology courses at a mid-sized southeastern university. A total of 464 students were recruited through an undergraduate subject pool and were granted psychology course credit for participation. Three responses were invalid or missing, yielding a final sample size of 461. The first few items on the survey prompted participants to report their gender, age, and race, thereby allowing for collecting demographic information on the sample. Participants were 377 female (81.8%) and 84 male (18.2%). The participants were predominantly Caucasian (83.7%), Asian-American (5.9%), African-American (5.2%), and Hispanic or Latino American (2.2%). The average student age was 19.3, ranging from 18.2 to 43.9 (SD = 2). The sample consisted of 281 freshmen (61%), 141 sophomores (30.6%), and 37 upper classmen (8%).

Measures

The full 50-item version of the Autism-Spectrum Quotient (AQ-50) was included as the first measure in the battery of other tests. This self-report measure was created to assess the degree of autistic traits among non-disabled adults with typical intelligence (Baron-Cohen et al., 2001). A subset of items from this questionnaire comprised the AQ-26 which is the focus of the current study (AQ-26 items are listed in Appendix A).The AQ-26 has never been administered in isolation from the other 24 items before; therefore

the full AQ-50 scale was administered in this study in order to avoid unexpected item ordering effects. Participants rated their agreement with each of the AQ-26 statements on a 4-point Likert-type scale ranging from "1" ("Definitely Disagree") to "4" ("Definitely Agree"). Eleven items were negatively worded and were reverse scored prior to conducting analyses. Higher scores on AQ-26 purportedly indicate a heavier autistic load (Baron-Cohen et al., 2001).

Analyses

Confirmatory Factor Analyses

The following analyses were conducted on the collected data to investigate the dimensionality of the AQ-26. Confirmatory Factor Analyses (CFA) were conducted to test the fit of the competing models. The CFAs tested the following models specified a priori based on previous research:

(1) Single-Factor Model (*Figure 1)*

(2) Five-Factor Model (Baron-Cohen et al., 2001) (*Figure 2)*

(3) Three-Factor Model (Austin, 2005; Hurst et al., 2007) (*Figure 3)*

(4) Higher-Order Factor Model (Hoekstra et al., 2008) (*Figure 4*)

(5) Four-Factor Model (Stewart & Austin, 2009) *(Figure 5)*

All CFAs were conducted using Mplus 5.2 with Weighted Least Squares Mean and Variance Adjusted (WLSMV) estimation method. This method was selected due to the categorical nature of 4-point Likert scale items. Such items result in categorical variables which cannot be normally distributed by definition (Finney & DiStefano, 2006). According to recent simulation studies, WLSMV yields asymptotically unbiased

parameter estimates as well as accurate fit indices (Flora & Curran, 2004). Given that the responses to the AQ-26 are categorical, the analysis of these items warrants the use of the WLSMV estimation.

Fit Indices. To evaluate the fit of these models, both absolute and incremental fit indices were interpreted. Whereas some fit indices are useful for evaluating simple model misfit (misspecified covariances between factors), others illuminate complex model misfit (misspecified relationships between items and factors). Therefore, both types are needed for evaluating overall model fit. The cut-off values of these indices recommended for the use with categorical data by Yu and Muthén (2005) were used.

The following absolute indices were examined. Mean- and variance-adjusted χ^2 is an absolute goodness-of-fit index assessing the degree of discrepancy between model-implied and observed covariance matrices. Non-significant values suggest that a proposed model is a plausible fit for the data. However, the χ^2 test is highly dependent on the sample size and tests a very stringent hypothesis. Therefore, other fit indices were examined. The root mean square error of approximation (RMSEA) is an absolute fit index sensitive to misspecified factor loadings. Lower values of RMSEA indicate better fit, with 0.05 suggested as a cut-off point (Yu & Muthén, 2005). As per incremental fit indices, comparative fit index (CFI) and Tucker-Lewis index (TLI) were used. These two indices are fairly sensitive to misspecified factor covariances and very sensitive to misspecified factor loadings (Hu & Bentler, 1998). A cutoff point of 0.96 is recommended for categorical variables. In addition, weighted root mean square residual (WRMR) was interpreted because it accounts for asymptotic variances, and is thus well-

suited for categorical indicators (Finney & DiStefano, 2006). Lower values of WRMR indicate better fit, with 1.0 suggested as a cutoff point (Yu & Muthén, 2005). Also, polychoric correlation residuals (the differences between observed and model-implied matrices) were examined in order to corroborate the results based on the fit indices.

Results of the model fit evaluation dictated the next steps in the analysis. If any of the models were found to fit the data and explain a substantial amount of variance in the items, these models weree compared and the best fitting and the most parsimonious model championed. Next, parameter estimates and factor reliabilities of the championed model were evaluated. In the case where none of the specified models fit the data, the next analytic step was to diagnose model misfit. The subsequent sections outline the stages of diagnosing model misfit.

Diagnosing Model Misfit. If none of the models adequately fit the data, the polychoric correlation residuals were examined in more detail because they highlight areas of local misfit. There is a polychoric correlation residual associated with each pair of items. Larger values indicate a greater degree of misspecification between the two items, with absolute values greater than |0.1| used as cutoff (Kline, 2005). Overall, large positive residuals indicate that a relationship between two indicators is underestimated; whereas large negative residuals suggest that a relationship between two items is overestimated.

Alternative methods of diagnosing model misfit were also employed, as needed. More specifically, exploratory factor analysis was conducted in order to diagnose model misfit and explore alternative solutions.

Comparing Two Scoring Schemes

As explicated earlier, two scoring schemes are currently used by different researchers to score the AQ-26. The first one is the 2-point scoring scheme, recommended by Baron-Cohen et al. (2001). This scoring scheme results in dichotomous variables. The second scoring theme commonly used by researchers retains the 4 points of a Likert scale, resulting in categorical variables. In order to test the equivalency of these two scoring schemes, CFA analyses outlined above were conducted twice using these two different scoring schemes. Comparing the results of these two analyses informed us as to whether these two scoring schemes are interchangeable.

CHAPTER 4

Results

Planned Data Analysis

Data analysis unfolded in the following phases. First, five theoretically specified models were fit to the data using CFA for categorical data. The CFA results were analyzed in two conditions, using the 4-point and the 2-point data, thus allowing us to assess the comparability of the two scoring schemes. Next, model misfit was analyzed using the three-factor model. Second, exploratory factor analysis (EFA) was conducted in order to identify poorly functioning items and explore naturally emerging factors.

Data Screening

After deleting three cases whose response strings consisted of excessive missing or invalid data, the final sample size was 461. Due to the forced response option activated during administration of the AQ-26, all 461 participants responded to all of the items. Therefore, none of the items were missing and no response patterns appeared to be invalid (i.e., there was no reason to suspect random response patterns).

The data were screened for multivariate outliers by calculating Mahalanobis distance for each student. Upon examination of these values, four cases were identified as potential outliers. However, a closer look at the pattern of these responses did not indicate that these students followed a response pattern indicative of the case being from a different population. Due to the low number of potential outliers and lack of suspicious response patterns, it was not deemed necessary to alter the data by removing these cases.

Since the variables under current analysis are comprised of discrete categories, they cannot be assumed continuous and normally distributed (Finney & DiStefano,

2006). This is the case regardless of which scoring scheme is used: 4-point or 2-point. Due to the normality assumption being not pertinent for categorical data, skewness and kurtosis were not reported. Also, measures of central tendency and standard deviations are not informative given the categorical nature of the data. Table 3 consists of response frequencies percentages for both 4-point and 2-point datasets. In the 4-point data, all categories had at least some responses in them, suggesting that sparse data was not an issue. It follows that the 2-point data also had data values in each category. The response category with the highest percentage of responses is bolded for each item. Although all items had responses in each category, some items were very unbalanced. For example, over 80% of respondents slightly or definitely agreed to the following items: 11, 12, 15, 44, 34, 38, 40, 44, 47, 50.

Pearson product correlations, which are commonly used to model relationships between observed continuous variables, cannot be used to represent the relationships between categorical variables (Brown, 2006). Appropriate correlations are needed to represent the linear relationships between the latent continuous variables, which are assumed to underlie the observed categorical variables (Flora & Curran, 2004). For this purpose, polychoric correlations are used with the 4-point data and tetrachoric correlations, which are a special case of the polychoric correlations, are used for the 2-point data. Table 2 presents both correlation matrices. Overall, the strength of the correlations was low in both conditions. Specifically, 74% of the polychoric correlations were below |0.2| and 67% of the tetrachorics were below |0.2|. Although Pearson product correlations decrease in magnitude when going from 4-point to 2-point, this is not the

case with polychoric and tetrachoric correlations. The tetrachorics were not systematically lower than polychorics. In fact, for 56% of correlations, tetrachorics were larger than polychorics (with the average difference being 0.06), and for 43% of correlations, polychorics were larger than tetrachorics (with the average difference being 0.04).

Confirmatory Factor Analyses

Confirmatory factor analyses (CFAs) using WLSMV were conducted in MPlus 5.2. Table 4 presents the results of two sets of CFA analyses: one conducted using the 4-point data and another conducted using the 2-point data. The following sections describe the fit of each model with particular attention paid to differences in model fit across the 4- and 2-point conditions. It is of interest to compare the fit of models in the 4-point and 2-point conditions in order to understand the effects of collapsing 4-category response scale into two categories.

Single-Factor Model (Figure 1). The first model fit to the data was the single-factor uni-dimensional model. If this model fit the data, it would be appropriate to calculate a total score and declare that the instrument measures one trait or construct. However, the single-factor model failed to yield adequate fit in both 4-point and 2-point conditions.

The change in fit indices was not consistent across the 4-point and 2-point conditions. For instance, absolute fit indices were more favorable in the 2-point condition (χ^2, RMSEA, WRMR) whereas incremental fit indices were less favorable (CFI, TLI).

Possible reasons for the discrepancy in how fit indices change across conditions are provided in the discussion.

Five-Factor Model (Figure 2). The next model fit to the data was the five-factor model which reflects the original five subscales created by Baron-Cohen et al. (2001). In the 4-point condition, the five-factor model did not provide adequate model-data fit, as indicated by all the fit indices. In the 2-point condition, the five-factor model failed to yield an admissible solution. Specifically, the factor covariance matrix was non-positive definite. Two relatively straightforward reasons why a factor covariance matrix may be non-positive definite include the presence of near zero factor variances or the presence of factor correlations near one. Both reasons were present in the 2-point condition. Although the estimated parameters cannot be meaningfully interpreted when a solution is inadmissible, they can provide insight as to why estimation of the model was problematic. There were high correlations between the factors *Attention to Switching* and *Attention to Detail* ($r = 0.946$) and between the factors *Social Skill* and *Communication* ($r = 0.903$). These results indicate that these factors are not truly distinct from each other and that the five-factor model might be "over-factored" (Brown, 2006). Interestingly, in the 4-point condition the correlation between *Attention to Switching* and *Attention to Detail* was far lower ($r = 0.467$) and the correlation between *Social Skill* and *Communication* was similar in value ($r = 0.892$), but not problematic for proper estimation. In addition, the variance of the *Imagination* factor was close to 0 in the 2-point condition (0.095). The variance of the *Imagination* factor was similar in the 4-point condition (0.197), but obviously not low enough to impede proper estimation. Because of

the presence of near perfect correlations between factors and a low factor variance, the five-factor model yielded an inadmissible solution in the 2-point condition and thus the results of this analysis cannot be interpreted.

Three-Factor Model (Figure 3). The three-factor model converged to an admissible solution in both the 4-point and 2-point conditions. Let us first consider the fit indices for the 4-point condition. Unfortunately, none of the fit indices approached the cutoff point indicating that the three-factor model is not a plausible model for the data. Although fit indices favored the simpler three-factor model over the more complex five-factor model, neither model provided satisfactory fit to the data.

The fit indices for the three-factor model in the 2-point condition also departed from the cutoffs, indicating that the three-factor model did not fit the data in either condition. As with the one-factor results, the change in fit indices across the two conditions was not consistent. The incremental fit indices (CFI, TLI) were less favorable in the 2-point condition and the absolute fit indices (χ^2, RMSEA, WRMR) were more favorable, almost to the extent that they approach the criteria for well-fitting models. Again, the possible reasons for the inconsistent change in fit indices across the 4-point and 2-point conditions are provided in the discussion.

Higher-Order Model (Figure 4). The higher-order model did not converge in either condition. Recall that the higher order model specified the higher-order factor of *Social Interaction* to subsume the following factors out of the five-factor model: *Social Skill, Attention Switching, Communication,* and *Imagination.* The remaining factor *Attention to Detail* stands independently in this model and is correlated with the higher-

order factor *Social Interaction.* The reason why the higher-order model failed to converge most likely lies in the low inter-factor correlations of the factors subsumed under *Social Interaction.* Specifically, the inter-factor correlations ranged from 0.19 to 0.89 in the 4-point condition with the five factor model, with the majority of these correlations being below 0.35. The higher-order model most likely did not converge because these inter-factor correlations were not high enough to be grouped under a single hierarchical model.

Four-Factor Model (Figure 5). Similar to both the five- and three-factor models, the four-factor model failed to provide adequate fit to the data in the 4-point condition. None of the fit indices surpassed the cutoff criteria. Although the fit indices for the four-factor model were marginally more favorable than those for the three-factor model, neither model fit the data.

As for the 2-point condition, the four-factor model failed to converge to an admissible solution due to a Heywood case. The problem had to do with item 50, which has a standardized loading above 1.00 (1.858) and accordingly, a R^2 value exceeding the maximum possible value of 1. Although the standardized loading for this item was high in the 4-point condition (0.973), it was not high enough to elicit problems with estimation. Because the presence of a Heywood case makes a solution inadmissible, the results of the 2-point condition are uninterpretable.

Which model is the best fitting? In the absolute sense, none of the tested models produced good model-data fit in either condition. It follows that the parameter estimates based on these models should not be interpreted (Brown, 2006; Kline, 2005). Given the lack of fit of any of the models, the next logical step is to analyze model misfit. However,

we need to choose the model that best approximates the observed data and thus allows us to accurately pinpoint the areas of local misfit.

Strictly speaking, the four-factor model provided the best fit in the 4-point condition, compared to the other models. However, closer examination of the fit indices reveals that the four-factor model fits only marginally better than the three-factor model. Another reason for looking at the three-factor model is that it is more parsimonious and differs only slightly from the four-factor model. Looking at the composition of a four-factor model (see Table 1), it becomes evident that the fourth factor specified in this model, *Imagination,* consists of only two items: 40 and 50. This calls into question whether an *Imagination* factor is really well-represented by these two items. Further, examining the content of these two items reveals that these two items are very similar in wording; they both talk about "playing games involving pretending with other children". Given this similarity, it might be that the common variance shared by these two items is simply due to a method factor, not to a valid psychological construct labeled *Imagination.* This suspicion is further substantiated by the problematic properties of item 50 in the 2-point condition. Furthermore, the three-factor model has garnered the most empirical and theoretical support in previous research on AQ (Austin, 2005; Hurst et al., 2007). For all of these reasons, the three-factor model will be the focus of subsequent analyses. In these subsequent analyses, model misfit in the three-factor model will be examined.

Analyzing Model Misfit

As justified above, the three-factor model was deemed the best fitting out of all the models considered in this analysis, and is thus the most interpretable. However, it still

does not fit the data, as illustrated by the fit indices in both the 4-point and 2-point conditions. Nonetheless, it is the most appropriate model for identifying areas of local misfit. Following Brown's (2006) recommendations for analyzing local misfit, the next section will focus on analyzing local misfit by examining the correlation residuals.

Correlation Residuals

The fit indices are informative in that they encapsulate the overall match between observed and model-implied correlation matrices. Although the fit indices provide a good summative measure of global model fit (how well the model reproduces reality overall), they mask areas of local misfit (specific misspecified relationships). Fortunately, the residual matrix provides information useful for diagnosing local misfit. The residuals in the matrix are simply the differences between the observed correlations and the model-implied correlations. The residual matrix contains a residual for each pair of items, with large positive values indicating that the model underestimates the relationship between items and large negative values indicating that the model overestimates the relationship (Brown, 2006). In the current analysis, correlation residuals are analyzed. As a rule of thumb in SEM with continuous data, absolute values of standardized residuals greater than 0.1 are considered to suggest a poorly specified relationship between two continuous indicators; the same cut-off point can be used for correlation residuals obtained from the analysis of the categorical items (Kline, 2005). Next, we provide a descriptive summary of the correlation residuals in the 4-point and 2-point conditions.

Correlation residuals in the 4-point condition. Out of 325 possible correlation residuals in the 4-point condition, 81 (nearly 25%) were above |0.1| indicating nontrivial

differences between model-implied and observed correlation matrices. The majority of misspecified relationships were underestimated, as indicated by 44 (13.5%) residuals above 0.1. However, a substantial number of misspecified relationships were also overestimated, as indicated by 34 (10.5%) residuals below -0.1. Sixteen residuals (4.9%) exceeded |0.2|, and three (0.92%) were above 0.3. Given the large number of residuals exceeding the cutoff point of |0.1| it was not feasible to easily identify a pattern of misfit.

Correlation residuals in the 2-point condition. Out of 325 possible correlation residuals in the 2-point condition, 117 (36%) were above |0.1|. The pattern of results was similar to that in the 4-point condition, with the majority of misspecified relationships (66, 20.3%) being underestimated and some being overestimated (51, 15.7%). Twenty eight residuals (8.6%) exceeded |0.2|, and six (1.8%) were above 0.3. Again, identifying a pattern of misfit is not feasible due to a high number of correlation residuals. Notably, the correlation residuals were somewhat larger in the 2-point condition than they were in the 4-point condition, even though the fit indices in the 2-point condition looked slightly better.

Severely misspecified relationships. There were so many correlation residuals surpassing the cutoff point, that it was not feasible to identify a pattern of misfit. However, it is worthwhile to consider the items whose relationships were grossly misspecified (residual above |0.3|). Table 5 contains the items whose relationships were severely underestimated by the three-factor model, the factor each item was supposed to load on, value of the correlation residual and possible reasons for misfit. Underestimated

relationship between two items indicates that they share common variance beyond what can be explained by their common factor.

Overall, the investigation of misfit in the three-factor model revealed that misfit is substantial and pervasive in both 4-point and 2-point conditions. Absolute fit indices suggested that the three-factor model did not yield adequate global fit to the data. Correlation residuals indicated that the local misfit is also substantial. In fact, a high number of correlation residuals prevented us from drawing meaningful conclusions regarding the overall pattern of misfit. However, severely misspecified item relationships were noted, along with the potential reasons causing the misspecification. Overall, the vast amount of correlation residuals led us to conclude that (a) the three-factor model does not adequately approximate the data and (b) identifying a pattern of misfit is not feasible in the restrictive mode of CFA. As a result, the next logical step was to switch to a more exploratory mode of psychometric investigation. The next sections present the results of the EFA analysis.

Exploratory Factor Analysis

Exploratory factor analysis (EFA) was conducted in order to identify poorly functioning items and explore other plausible models underlying the AQ-26 (Brown, 2006). The EFA was conducted in Mplus 5.2 using the polychoric correlation matrix. The overall pattern of results was consistent across the 4-point and 2-point conditions, so it was considered most informative to interpret the results from the 4-point condition. EFA is an exploratory technique that allows factor solutions to emerge freely from the data. Similar to CFA, EFA yields fit indices for each model thus allowing for gauging model-

data fit. In the current analysis, 8 factor solutions were allowed to emerge. Table 6 presents fit indices for the 8 factors yielded by EFA on the AQ-26 in the 4-point condition. Expectedly, fit indices improved as the models increased in complexity. It was decided to more thoroughly examine the solutions consisting of 3 to 6 factors. The three factor solution was chosen as the most parsimonious model to further examine because several previous exploratory studies have advocated for a three-factor solution. The six-factor solution was chosen as the most complex model to further examine because acceptable fit for this model was obtained for all fit indices. Notably, these 3 to 6 factor models emerging in the current EFA were empirically derived and thus did not align with any theoretical conceptualization of AS. Instead, they were used to explore the dimensionality of the AQ-26. In examining the results of these four exploratory models (3-factor, 4-factor, 5-factor and 6-factor), attention was devoted to item quality, consistency of model configurations across solutions, and interpretability of factors.

The factor pattern coefficients for models 3-6 and the correlations among factors in each solution are presented in Table 7. Items were deemed to "load" on a factor if their pattern coefficients were larger than |.30|. Factors will be referred to according to their factor number in the six-factor solution. Presentation of the EFA results will unfold in the following phases. First, the most interpretable and stable factors that consistently emerged across solutions are examined. Second, factors that are less interpretable or unstable are considered. Third, we compare and contrast how these exploratory factor structures compare to those championed in previous studies.

Interpretable and Stable Factors

Factor 1. The most stable factor, consistently emerging across solutions, consisted of 6 items all appearing to tap into one's inclination to pay attention to patterns, details and numerical information. The item with the strongest loading (pattern coefficient) is 23 "I notice patterns in things all the time". Other items loading on this factor pertained to fascination or attention to numerical data of some sort (e.g., numbers, dates) or details. Item 5 "I often notice small sounds when others do not" has the lowest pattern coefficient, indicating its weak relationship to this factor. However, this item did not "travel" to different factors across solutions, suggesting that it is pertinent to this factor, but not as salient as the other items on this factor. Perhaps noticing small sounds is not exactly the same as the content being addressed by the other items.

Factor 5. The next factor that appeared to consistently emerge across solutions is factor 5 which consisted of two items, one asking about how disturbing it would be to have one's daily routine interrupted and the other inquiring about the need to carefully plan one's daily activities. This factor eventually merged with factor 3 in the three-factor solution, however the content of items on factor 5 and factor 3 (described below) appear quite distinct. Although the similarity between the two items on factor 5 is evident, two items are rarely considered enough to cover the breadth of a factor. If the need for control over one's routines and activities is deemed an important criterion for assessing autistic traits in the general population, then more items may need to be written to tap into this factor.

Uninterpretable or Unstable Factors

Factor 2. This factor consisted of items tending to deal with comfort in social interactions and preferences for social situations. The two items that loaded the highest on this factor across solutions had to do with enjoying or being good at social chit-chat. In the six factor solution, ten items had pattern coefficients above 0.30 on factor 2. Although there are several items that consistently load on this factor across solutions (38, 17, 47, 11, 22, 15), this factor was designated as being inconsistent because many other items cross-loaded on other factors across solutions. Cross loadings indicate that items are factorially complex, tapping into more than one dimension. For instance, item 34 "I enjoy doing things spontaneously" also loads on factor 5 in addition to factor 2 in many solutions. It makes sense that this item would load on factor 5, which pertains to one's need for control over their routines and activities. It is harder to argue that this item belongs on factor 2, which deals with sociability. Despite the rationale for why the item should load on one or the other factor, empirically the item relates to both factors. Unless there is a strong justification for the cross-loading, items that cross-load should be revised, substituted, or omitted altogether are most likely to introduce misfit in subsequent CFAs. Overall, there are some salvageable items on this factor, but substantial work is needed to improve several of the cross-loading items.

Factor 3. Factor 3 consisted of five items pertaining to understanding others' intentions and picking up on non-verbal social cues (45, 35, 20). There are a few issues associated with this factor. First, the pattern coefficients are quite low (ranging from 0.31 to 0.53 in the six-factor solution), indicating that these items are not *strong* indicators of this factor. Also, one item (39) cross-loads in the six-factor solution and seems different

in content than the remainder of the items ("People often tell me that I keep going on and on about the same thing"). Another item, item 37 "If there is an interruption, I can switch back to what I was doing very quickly" also does not seem meaningfully related to other items. This item also fails to load on any factor in the other solutions. Therefore, these two items are candidates for exclusion or revision.

Factor 6. Factor 6 has two items that inquire about playing pretend games with children. This 2-item factor is found in the five- and six-factor solutions, but disappears in the remaining solutions. In the remaining solutions, the items fail to load on any factor. One recommendation for this factor is similar to that provided with Factor 5. That is, more items would need to be created for this factor because it is doubtful that two items sufficiently represent the breadth of the construct being assessed. However, it is difficult to argue that the construct "playing pretend games with children" is a factor necessary for measuring autistic traits in the general population. The authors created these items to tap into the more general construct "imagination", but the results suggest that instead, these two items are tapping into their own limited construct that does not seem to be related to the other imagination items (or any other items on this scale). For this reason, it is suggested that the inclusion of these items on the scale be reconsidered.

Factor 4. Item 7 "Other people frequently tell me that what I've said is impolite, even though I think it is polite" is subject for omission. It was the only item loading solely on factor 4 in the six-factor solution. The other items loading on factor 4 in this solution seemed to have little in common with item 7 and also cross-loaded on other factors. Factor 4 merged with factor 1 in more parsimonious solutions, but no sound

rationale can be provided for the union of item 7 with the other factor 1 items. The content of this item pertains to politeness, which seems unrelated to any of the other factors.

Comparing to Theoretical Models

When comparing the solutions from the EFA to those offered by other researchers, certain similarities and differences become evident. First, the five-factor structure originally posited by Baron-Cohen et al. (2001) yielded very poor fit in a CFA, and did not emerge in the EFA, further corroborating previous findings (e.g., Austin, 2005; Hurst et al., 2007). The higher-order model championed by Hoekstra et al. (2008) failed to converge in a CFA due to the low inter-factor correlations in this sample. This is not surprising given that the higher-order model is based on the Baron-Cohen's five factor model, which has little empirical support.

The factors emerging in the EFA most strongly resemble the three-factor solution supported by Austin (2005) and Hurst et al. (2007) and the four-factor solution advocated for by Stewart and Austin (2009). Recall that the four-factor solution of Stewart and Austin (2009) only differed from the three-factor solution found in previous studies in that items 40 and 50 loaded on their own factor (titled *Imagination*) as opposed to loading on the *Socials Skills* factor. Table 8 compares the three-factor solution advocated by Austin (2005) and Hurst et al. (2007) with the six-factor EFA solution. In the current EFA analyses, factor 1 appeared to tap into one's inclination toward patterns and details, and consisted of the six items that also loaded on *Details/Patterns* factor in previous studies. However, there were two items loading on the *Details/Patterns* factor that were

not found to load on factor 1 in this study. These items are: 25 ("It does not upset me if my daily routine is disturbed") and 43 ("I like to plan any activities I participate in carefully"). These two items actually loaded on their own factor in EFA – Factor 5, a factor that appears to tap into one's inclination for routines. The fact that the items on factors 1 and 5 loaded on the same factor *Details/Patterns* in previous studies raises the question of how distinct Factors 1 and 5 are from one another. Looking at the content of these two factors, it appears that they address distinct characteristics: Factor 1 pertains to details or patterns, whereas factor 2 pertains to the need for routine. Furthermore, there is no empirical indication that the two factors should be combined as these two factors never merged in any of the EFA solutions. As well, the correlation between these factors in the EFAs was always small in size.

The next factor emerging in the current solution and also found in previous studies is *Social Skills.* The composition of Factor 2 in current EFA analysis is identical to that found by Stewart and Austin (2009). However, several items from this factor had split loadings in the current analysis: 44, 26, 13, and 34. These items are likely to cause misfit in a CFA and should either be revised or omitted, to ensure clean and simple structure. Notably, the two items included under this factor in Austin (2005) study (40 and 50) loaded on its own factor both in Stewart and Austin (2009) and in the present study (Factor 6). Stewart and Austin (2009) labeled this factor *Imagination,* but it appears to tap exclusively into playing pretend with children. Again, the factor with just two items is subject for revision or omission for reasons outlined above.

The factor labeled *Communication* by Austin (2005) and *Understanding Others/Communication* by Stewart and Austin (2009) also emerged in the current EFA as Factor 3. The composition of this factor is very similar to that found in the previous studies, with the following items loading on it: 45, 35, 20, 39, and 37. However, item 7 "Other people frequently tell me that what I've said is impolite, even though I think it is polite" was assigned to this factor in prior studies, but loaded on its own factor in the current EFA. Looking at the content of this item, it appears that this item deals more directly with politeness than understanding other people's intentions, like other items on this factor. Also, item 39 "People often tell me that I keep going on and on about the same thing" cross-loaded on factors 3 and 4 in a 6-factor solution. Item 37 "If there is an interruption, I can switch back to what I was doing very quickly" failed to load on any factor in other EFA solutions. Therefore, items 7, 37, and 39 are likely causing misfit in the CFA model.

CHAPTER 5

Discussion

Given the increased interest in self-report measures of autistic traits among typical adults, it is imperative to garnish validity evidence for such instruments. The first and the most widely used measure of Asperger Syndrome, the AQ, warrants rigorous psychometric investigation. The short version of the instrument, the AQ-26, has been understudied despite its promise as a more parsimonious measure. Although the research investigating the functionality of the AQ-26 has been started, the results of the studies thus far are inconclusive and prompt more questions than answers. Specifically, the factor structure underlying the responses to the AQ-26 is ambiguous, although multiple competing models have been proposed. Also, evidence regarding the exchangeability of the two different scoring schemes used (4-point and 2-point) is lacking. Furthermore, the methods used in most previous studies are not the most suitable for categorical data or psychometrically sound. Therefore, proper psychometric examination of the AQ-26 is needed so that researchers interested in using this scale can have more confidence in the inferences made based using its scores. Following Benson's (1998) strong program of construct validation, the current study focused on the structural stage of instrument development. More specifically, the results of this empirical investigation will be discussed in the following phases. First, the dimensionality of the AQ-26 is discussed in light of the current results and prior hypotheses. As part of this section, areas in need of special attention are highlighted. Second, we compare and contrast the two scoring schemes and provide recommendations on the number of scale categories. Finally, the

limitations of the current study are listed along with the implications these results might have for the future of the AQ-26 and AQ-50.

Dimensionality

CFAs and Model Misfit

Based on the previous research conducted on the full and short versions of the AQ, we identified five models theorized to underlie the responses to the AQ-26. Unfortunately, the results of CFAs revealed that none of the models produced adequate model-data fit. Furthermore, investigation of the correlation residuals to identify areas of local model misfit revealed that the misfit was pervasive. The next step was to switch into a more exploratory mode of analysis in order to effectively convey causes of misfit and to identify other plausible models for the AQ-26.

EFAs

During the exploratory phase of this study a variety of different EFA solutions were inspected and compared with those of previous EFA studies. Many similarities were found between the current study's EFA results and Austin's (2005) results, providing some support for the three-factor model first proposed by Austin (2005) and partially supported by Hurst et al. (2007). Like Austin (2005) and Hurst et al. (2007), three factors - Factors 1, 2, and 3 - emerged from the current study, tapping into *Details/Patterns, Social Skills*, and *Communication/Mindreading,* respectively. Although our results somewhat correspond with these previous studies, the ways in which they depart have important implications for the AQ-26. Therefore, there are several important caveats pertaining to this factor solution that need to be noted and are discussed for each factor.

The closest correspondence in results was in Factor 1, the *Details/Patterns* factor. Six items that loaded on the *Details/Patterns* factor (Factor 1) in the current study also loaded on this factor in other studies. However, two additional items (25 & 43) that were affiliated with this factor in previous research formed their own factor (Factor 5) in the current study. Because Factor 5 had only minor correlations with the *Details/Patterns* factor (Factor 1) and because these factors never merged at any point in the EFA solutions, there is no indication that these two items should ever be considered part of the *Details/Patterns* factor. As well, the fact that the content of these two items appears quite distinct from those of the *Details/Patterns* factor is another argument for their separation.

Factor 2, the *Social Skills* factor, was more problematic than Factor 1. Although Factor 2 in the current study shared many items affiliated with the *Social Skills* factor in other studies, several items were problematic because of their consistent cross-loadings. Also, two items (40 & 50) assigned to the *Social Skills* factor in Austin (2005) loaded on their own factor (Factor 6) in the current study, just as they did in Stewart and Austin (2008). Because Factor 6 had minor correlations with the *Social Skills* factor (Factor 2) and because these factors never merged at any point, there is no indication that these two items should ever be considered part of the *Social Skills* factor. As well, the fact that the content of these two items appears quite distinct from those of the *Social Skills* factor is another argument for their separation. In addition, item 37, which was considered part of the Social Skills factor in previous studies, did not strongly affiliate with this factor (or any other interpretable, stable factor) in our solutions.

The most problematic factor of all was Factor 3, the *Communication/Mindreading* factor. This factor was the most problematic not only because the items loading on it had low loadings, but also because of a consistent cross-loading item (39). As well, previous research affiliated item 7 with this factor, although our study provided little indication that this item belonged to this factor or any other interpretable, stable factor in our solutions.

The results of the current study indicate potential for the *Details/Patterns, Social Skills*, and *Communication/Mindreading* factors and provides direction for further refinement of these factors. Problematic items (e.g., 37 and 7), items with cross-loadings, or items with low loadings should either be omitted or revised. Careful consideration should be given to those items on Factors 5 and 6. If the content of these factors is considered essential to assessing autistic traits in the general population, adding these factors to the three-factor solution to create a five-factor solution is plausible. However, additional items would need to be acquired for these factors before consideration of this five-factor solution. Given that there is a 50-item measure, it would be prudent to consider other items on the AQ-50 for inclusion on the short form.

Although it was encouraging to see similarities between the current EFA results and those of Austin (2005) and Stewart and Austin (2008), the results departed in important ways. Given the discrepancy between the findings, one might ask which study's results are most trustworthy? It could be argued that the current study's results are the most trustworthy because the analysis: (a) focuses specifically on the AQ-26 and (b) uses correlations and estimation techniques more suitable for the categorical nature of the

responses. It will also be argued later in the discussion that the current study's findings are preferable because the 4-point, rather than 2-point, scoring scheme was utilized.

EFA after CFA

The EFA analyses in the current study provided quite a bit of insight into the structural validity of the AQ-26. Not only were we able to compare and contrast our EFA results with those of other studies, but we were also able to corroborate and provide further insight into the misfit of the CFA models. For instance, many of the problematic items identified when looking at the correlation residuals were also identified as troublesome in the EFA solutions. The EFA analyses were also beneficial in identifying causes of misfit. Although it was difficult to summarize misfit in the three-factor CFA model because of the large number of sizeable correlation residuals, it was relatively easy to understand why the this model did not fit after inspection of the EFA solutions. The EFA solutions also provided insight into why the other CFA models did not fit the data, thus avoiding the tedious and difficult task of summarizing each CFA model's correlation residuals. It is recommended that future researchers pursue EFA after failing to acquire fit with their CFA models, particularly if the instrument is in the beginning stages of development.

Future of the AQ-26

It is very important to note that the original five-factor model posited by Baron-Cohen et al. (2001) garnered no support in the current study. Also, other psychometric studies of the AQ failed to support this model (e.g., Hurst et al., 2007; Stewart & Austin, 2009). Similarly, the unidimensional model was not supported in the current or previous

studies, providing further evidence that the AQ does not measure a single construct (e.g., Hoekstra et al., 2008). There are very important practical implications to these consistent findings regarding the implausibility of the 1- and 5-factor models. Namely, it is inappropriate to score the AQ-26 using five subscales or the total score. However, a question of how to appropriately score the AQ-26 remains unanswered. Although the results of the current study provide insight into how the items or scoring models could be revised, it does not strongly advocate for a particular scoring scheme for the current AQ-26.

The summary of the AQ-26 functionality provided herein allows us to draw important conclusions about the future use of the AQ-26. First, the AQ-26 should be used cautiously in its current form because little support has been gathered for the structural validity of the scale and it is unclear how to score it. In order to improve the scale, future research efforts can focus on obtaining a clean and simple factor structure by developing better items. Beginning with the three-factor solution of Austin (2005) and considering how our EFA solutions correspond with this solution is a first step.

Before substantial modifications are made to the AQ-26, it is important to question whether this particular short form warrants more attention. It is possible that the 26-item scale is not the best selection of items for a short form given that these items were selected based on PCA for continuous data (Austin, 2005). Therefore, it might be useful to investigate the AQ-50 using appropriate methodology and create a short version of the scale based on these analyses. Nonetheless, the current study offers insight about assessing autistic traits in the general population. The next sections focus on interpreting

the results of the current study in light of the theoretical understanding of AS. First, we will consider the diagnostic criteria stipulated by the DSM-IV. Next, we will consider additional criteria commonly identified by clinicians to characterize AS.

Back to Theory

Social Impairment

Recall that AS is a primarily social disorder, characterized by impairments in social skills. DSM-IV currently prescribes that at least two of the following social impairments must be present for diagnosis: (a) *Nonverbal Behaviors,* (b) *Lack of Peer Relationships*, (c) *Lack of Social Spontaneity*, (d) *Lack of Social Reciprocity*. Across psychometric studies of the AQ, the dimension labeled either *Social Skills* or *Socialness* emerges consistently. However, there are a few issues with this dimension. First, this dimension is treated as a single, congeneric factor in most of the studies, whereas it is fragmented into four sub-components in the DSM-IV criteria. The exception to this is the higher-order model championed by Hoekstra et al. (2008). In that model, a higher order factor *Social Interaction* subsumes four lower-order factors. Unfortunately, this model is unlikely to underlie the responses to the AQ-26 because it is based on the five-dimensional model of Baron-Cohen et al. (2001) which has failed to garner support in this and other studies. Furthermore, the analysis of model misfit in the present study revealed multiple areas of concern associated with the *Social Skills* factor. Although the same ten items loaded on this factor as they did in Austin (2005) and Hurst et al. (2007), several items had split loadings, indicating that they are multi-factorial indicators of more than one factor. These items are: 44 (“I enjoy social occasions”), 26 (“I frequently find

that I don't know how to keep a conversation going"), 13 ("I would rather go to a library than a party**"**), and 34 ("I enjoy doing things spontaneously"). So, these split loadings indicate that *Socialness* might be multi-dimensional construct. In conjunction with the theoretical conceptualization of Social Impairment as a multi-faceted construct, these findings are quite telling. This hypothesis is supported by early research on assessing basic social skills among typical adults conducted by Riggio (1986). In that work, six facets of social skills were identified: emotional expressivity, emotional sensitivity, social expressivity, and social sensitivity, emotional control, and social control. Given that social functioning is so important to classifying AS, it may be that *Socialness* factor warrants a more elaborative treatment by this scale, perhaps including several factors tapping into the different dimensions of social ability.

Another concern about the *Social Skills* factor has to do with the differences between measuring preferences and measuring abilities. Looking at the content of the items, it is clear that some items get at the preferences because they start with "I enjoy…" or "I would rather…". Items were written in this manner on purpose, guided by the presumption that individuals are more likely to accurately report their preferences as opposed to providing an accurate judgment on their own behavior (Baron-Cohen et al., 2001). Other items appear to assess one's own assessment of social competence. An example of this is item 22 "I find it hard to make new friends". However, the DSM-IV criteria clearly states that impairment in social interaction characterizes AS, not the locus of preference. Although the assumption here is that preferences are indicative of the abilities, this assumption lacks empirical evidence. Overall, it seems that the *Social Skills*

factor should receive more attention during future instrument development and validation.

Behavioral Stereotypy

It appears that the AQ-26 does not fully address another important diagnostic criterion for AS – behavioral stereotypy. As currently listed in the DSM-IV, this category subsumes: (a) *Restricted Interests*, (b) *Adherence to Routines*, (c) *Repetitive Motor Mannerisms*, and (d) *Preoccupation with Parts of Objects*. It appears that *Restricted Interests* does not align with any of the factors emerging in the previous or current studies. This might be a major pitfall of the scale, especially given that many researchers specializing in the autism spectrum disorders render behavioral stereotypy to be salient to AS (Attwood, 2007; Myles & Simpson, 2002). Future research efforts can focus on developing items tapping into this dimension. Interestingly, *Adherence to Routines* was not specified as a factor by the scale authors and it did not emerge as a dimension in any subsequent psychometric analyses, except for the present exploratory factor analysis. Herein, two items loaded on a factor labeled *Need for Routine*: 25 ("It does not upset me if my daily routine is disturbed.") and 43 ("I like to plan any activities I participate in carefully."). Factors consisting of only two items are unfavorable because two items are unlikely to fully cover the breadth of the construct. If this dimension is indeed theoretically important, more items should be written to tap into it. The next dimension, outlined in the DSM-IV is *Repetitive Motor Mannerisms*. Currently, this dimension is absent from the AQ, probably because it is difficult to assess via self-report. Nonetheless, an effort should be made to assess this behavioral tendency. The next

criterion *Preoccupation with Parts of Objects* might be related to the consistently emerging factor labeled *Details/Patterns.* However, theoretical justification for the equivalence of these two constructs is needed. Although diagnosis of the AS requires only one of the symptoms above to be present, it seems that a comprehensive screening scale should cover as many dimensions as possible.

Other Features of Asperger Syndrome

Recall that other typology, beyond the scope of impairments outlined in the DSM-IV, is known to be characteristic of AS. Some of these features are: impairments in communication (Attwood, 2007; Baron-Cohen et al., 2005), executive functioning (Foster, 2002; Schutte, 2003), sensory issues (Dunn et al., 2002), and motor issues (Smith, 2000). Aspects of communication ability are addressed in several investigations of the AQ, but the results across these studies are not conclusive. For example, Baron-Cohen et al. (2001) originally wrote items to tap into Communication but this factor did not emerge in the subsequent analyses. Instead, items pertaining to one's ability to work out other people's intentions loaded more consistently on a single factor labeled Mindreading by Austin (2005). Although the same factor emerged in the current study also, it had a few problematic items. Whether or not to improve this factor and include it in the scale depends on whether mindreading is considered an important aspect of autistic traits. Similarly, the answer to the question of whether or not "playing pretend with children" is pertinent to autistic traits will determine whether the factor tapping into this characteristic should be further refined and included in the scale. Analogously, theory should be consulted to determine whether such features of AS as executive functioning,

sensory issues, and motor issues ought to be addressed by the AQ. Given numerous theoretical conceptualizations of AS, one might argue that creating an instrument tapping solely into the core features of AS is more prudent than attempting to address every single aspect of AS. On the other hand, it can be argued that a sound instrument should be as thorough as possible and address most of the known aspects. This dilemma is not likely to be resolved before the scientific community comes to a consensus as to what exactly constitutes AS among adults.

Comparing Scoring Schemes

The effects of using two different scoring schemes on the results of CFA were examined. These two scoring schemes were: 4-point (consisting of 4 Likert scale categories) and 2-point (4 categories collapsed into two). Recall that originally, the authors of the AQ recommended the 2-point scoring scheme to facilitate easy scoring. The authors performed item analysis (by looking at the percentage of different groups scoring on each item) and concluded that "regarding the decision to score "slightly agree" and "definitely agree" responses using 1 point only, a reanalysis differentiating these in terms of 1 versus 2 points led to the same pattern of results overall" (Baron-Cohen et al., 2001, p. 11). However, Baron-Cohen and his colleagues did not conduct a factor analysis on the scale and did not test the comparability of these two scoring schemes on the results of factor analyses. Some of the subsequent psychometric studies of the AQ used the 2-point scheme (e.g. Hurst et al., 2007) and others used the 4-point scheme (e.g., Stewart & Austin, 2009; Hoekstra et al., 2008). Researchers opting to choose the 4-point scheme argue that four points capture more information, increase inter-item correlations, scale

reliability and validity coefficients (Stewart & Austin, 2009). However, no systematic analysis of the differences between the two scoring schemes has been conducted. In order to understand the effects of using 4-point versus 2-point scoring schemes, it is necessary to understand how the number of categories affect correlation matrices. First, let us consider the effects of dichotomizing variables on the Pearson product moment (PPM) correlations and resulting CFA fit indices.

Even though PPM correlations are not appropriate for the use with categorical data, it is informative to investigate the differences in the 4-point and 2-point conditions when PPMs are used. PPM correlations using 2-point data are weaker in magnitude than PPM correlations using 4-point data. This difference in the 4- and 2-point PPMs leads to important differences in the results of factor analyses using 4- and 2- point data. For instance, the loadings would be lower in the 2-point data than in the 4-point data. The model's χ^2 would be lower in the 2-point versus 4-point because lower correlations are easier to reproduce. For this same reason, the RMSEA and SRMR would look better in the 2-point condition. Incremental fit indices (CFI and TLI), which capture the degree of improvement in fit of the target model comparative to the baseline model[6], would be worse (lower) in the 2-point condition. This is due to the target model not being much of an improvement over the baseline model in 2-point data. In summary, factor loadings and comparative fit indices would look less favorable with 2-point data and absolute fit indices would look more favorable.

[6] The baseline model assumes zero population covariances among the observed variables (Kline, 2005).

Although this same pattern of change in the fit indices across the 4- and 2-point solutions was found in the current study when polychoric and tetrachorics correlation matrices are used, the reasons for these differences change. Recall that PPM correlations are not appropriate for categorical data because they fail to accurately capture the relationships between discrete categories (Brown, 2006). Instead, polychoric and tetrachorics correlation matrices should be used instead because they take into account floor and ceiling effects of the categorical data, are larger in magnitude, and yield more accurate fit indices (Brown, 2006). Comparing values of the polychorics (for the 4-point) and tetrachorics (for the 2-point) reveals that there are no systematic differences between these two coefficients. If anything, the tetrachorics are slightly stronger than the polychorics. Therefore, in this situation, the *magnitude* of the correlations is somewhat comparable.

However, the fit indices still differ across these two conditions in the same manner as they are expected to differ with PPMs. The reason for the discrepancy in fit indices across the 4- and 2-point conditions is attributable to the lower PPMs in the 2-point condition. However, given that tetrachorics are not systematically lower than polychorics, this same explanation does not hold for categorical data. Unfortunately, it is not clear why the fit indices differ across the two conditions with categorical data.

Regardless of why the fit indices were discrepant across conditions, the difference in the factor analytic results using the 4-point and 2-point scoring schemes has practical implications. Researchers analyzing the fit of a model might arrive at different conclusions depending on which scoring scheme is used. For instance, in the current

study the three-factor CFA model did not yield adequate solution in the 4-point condition, whereas in the 2-point condition this model provided close to adequate fit, at least according to some indices. To prevent confusion, it is important than one scoring scheme is used consistently across psychometric and substantive studies. Another practical implication of the scoring scheme utilized has to do with estimation issues. Many of the CFA models in the 2-point condition could not be interpreted because the solution was inadmissible. The same models that were problematic in terms of estimation in the 2-point condition converged without complications in the 4-point condition.

Based on the empirical results of the current study, our recommendation is to use the 4-point scoring scheme. The reasons for this recommendation are as follows. First, the 4-point scale captures more information than the 2-point. Second, it appears that the CFA conducted on the 2-point data yields more convergence problems. Moreover, the justification for using the 2-point scoring is rather weak. Thus, the 4-point scoring scheme is preferred over the 2-point scheme. Alternatively, it might be worthwhile to consider response scales consisting of more than 4 points so that the use of factor analytic techniques appropriate for continuous responses is a possibility for the AQ.

Limitations of the Current Study

There are a few important limitations that need to be noted. First, the current analysis focused on the short version of the scale – the AQ-26. Although the current study can be informative for future psychometric investigations of the AQ-50, the results cannot be directly generalized to the AQ-50. However, several of the findings in the current study raise questions about the full-length scale. Specifically, none of the

theoretically posited models fit the responses underlying the AQ-26, suggesting that these models are also unlikely to be plausible for the AQ-50. Despite our focus on the short-form, we hope that readers take our results into consideration when evaluating the structural validity or making revisions to the full-length scale.

Second, the measure in the current study was administered on a computer, whereas in other studies it was given as a paper-and-pencil test. The exchangeability of the administration modes has not been empirically established, so there might be mode-specific response patterns that would not have been found in the paper-and-pencil administration. Third, the current study is only the second psychometric investigation of the AQ on an American sample. American students might be interpreting the items differently from the British and Dutch students, thus limiting the generalizability across studies.

Directions for Future Research

Conducting think-alouds or comprehension checks might be fruitful ways to explore the AQ-26. Such qualitative inquires, conducted prior to the next factor analyses, can shed light on the thinking processes that individuals engage in as they are responding to the items (Erickson & Simon, 1993). For example, it might be that the items were phrased in British English and thus were understood differently by the American college students. For example, “social chit-chat” and “car number plates” are not commonly used in American English. Furthermore, pursuing measurement invariance studies on the scale can be helpful in establishing functional equivalence of the scale across nations. Another possible way to improve the AQ is to tailor it toward college students specifically. Recall

that students with AS are likely to struggle with the various aspects of university life, such as working in groups, attending to multiple tasks, and making friends. Given that most of the research on AQ thus far has been conducted using student populations, it seems logical to fine-tune the AQ to address the specific features of *students* with suspected AS.

Alternatively, the attention can be switched toward other existing self-report scales intended to measure AS traits among adults. Two of such measures exist: AASS and RAADS. Although these instruments are not nearly as popular as the AQ, they are nonetheless very promising and warrant more study.

Conclusion

In summary, the current study provided a thorough psychometric investigation of the short version of the Autism Spectrum Quotient – the AQ-26. The focus of this study remained on the structural stage of construct validation (Benson, 1998) and provided valuable insight into the dimensionality of the scale, as well as the functionality of two scoring schemes. Essentially, the results of this study failed to support the structural validity of the AQ-26, but did provide guidance for instrument revision. It is recommended that future psychometric investigations of both AQ-50 and AQ-26 use the guidance provided here along with a careful consideration of the theoretical conceptualizations of Asperger Syndrome to create a stronger instrument. The results of this study also indicate that the 2-point and 4-point scoring schemes are not interchangeable and that the 4-point scoring scheme is preferred as it captures more information. In summary, the current study provided evidence that substantial work is

needed to improve the psychometric properties of the AQ-26 and suggested possible directions for future research.

Table 1

Item-Factor Mapping of Different Models of the AQ-26 (N = 461)

Items	Higher-Order Factor (Hoekstra et al., 2008) Social Interaction					Three-Factor (Austin, 2005; Hurst et al., 2007)			Four-Factor (Stewart & Austin, 2009)			
	Social Skills	Attention Switching	Communi cation	Imaginatio n	Att to Detai	Social Skill	Details/ Patterns	Comm/ Mindread	Socialness	Patterns	Understan ding	Imagination
11 I find social situations easy.	V					V			V			
13 I would rather go to a library than a party.	V					V			V			
15 I find myself drawn more strongly to people than to things.	V					V			V			
22 I find it hard to make new friends.	V					V			V			
44 I enjoy social occasions.	V					V			V			
45 I find it difficult to work out people's intentions.	V							V			V	
47 I enjoy meeting new people.	V					V			V			
5 I often notice small sounds when others do not.		V					V			V		
6 I usually notice car number plates or similar strings of information.		V					V			V		
9 I am fascinated by dates.		V					V			V		
25 It does not upset me if my daily routine is disturbed.		V					V			V		
34 I enjoy doing things spontaneously.		V				V			V			
37 If there is an interruption, I can switch back to what I was doing very quickly.		V						V			V	
43 I like to plan any activities I participate in carefully.		V					V			V		
7 Other people frequently tell me that what I've said is impolite, even though I think it is polite.			V					V			V	
17 I enjoy social chit-chat.			V			V			V			
26 I frequently find that I don't know how to keep a conversation going.			V			V			V			
35 I am often the last to understand the point of a joke.			V					V			V	
38 I am good at social chit-chat.			V			V			V			
39 People often tell me that I keep going on and on about the same thing.			V					V			V	
20 When I'm reading a story, I find it difficult to work out the characters' intentions.				V				V			V	
40 When I was young, I used to enjoy playing games involving pretending with other children.				V		V						V
50 I find it very easy to play games with children that involve pretending.				V		V						V
12 I tend to notice details that others do not.					V		V			V		
19 I am fascinated by numbers.					V		V			V		
23 I notice patterns in things all the time.					V		V			V		

Note. The five-factor model (Baron-Cohen et al., 2001) is equivalent to the Higher-Order factor model, except that all five factors stand independently. Under the unidemensional model , all items are summed up.

Table 2

Correlations for the AQ-26 (N =461)

Item	5	6	7	9	12	13	19	20	22	23	26	35	39	43	45	38	11	44	17	47	40	15	34	50	25	37
5	--	0.30	0.12	0.13	0.36	0.09	0.11	-0.03	0.06	0.23	-0.02	-0.14	-0.05	0.13	-0.13	0.00	0.00	0.06	0.07	0.04	-0.07	0.11	0.00	0.03	0.12	-0.08
6	0.31	--	0.29	0.37	0.47	0.22	0.41	-0.09	0.16	0.48	0.03	-0.22	-0.05	0.01	0.05	-0.01	0.01	0.24	-0.02	-0.02	-0.13	0.16	0.13	0.08	-0.01	-0.14
7	0.15	0.21	--	0.21	0.26	0.17	0.27	0.09	0.36	0.09	0.23	0.00	0.29	0.01	0.17	-0.07	-0.07	0.03	-0.06	-0.01	-0.03	-0.02	-0.16	0.05	0.00	-0.10
9	0.15	0.35	0.20	--	0.40	0.24	0.42	0.11	0.05	0.36	-0.01	-0.10	0.16	0.10	0.06	-0.02	-0.08	0.12	0.13	-0.07	-0.11	0.04	-0.04	-0.24	-0.16	-0.34
12	0.29	0.32	0.14	0.22	--	0.04	0.19	0.03	-0.07	0.48	-0.02	-0.25	0.07	0.02	-0.09	-0.17	-0.21	-0.09	0.04	-0.11	0.15	-0.05	-0.16	0.13	0.00	-0.06
13	0.07	0.18	0.12	0.26	0.07	--	0.19	0.02	0.40	0.22	0.24	0.16	0.09	0.19	0.17	0.32	0.38	0.67	0.38	0.36	0.15	0.39	0.39	-0.04	0.17	-0.05
19	0.08	0.38	0.14	0.43	0.18	0.15	--	0.16	0.18	0.46	-0.01	-0.10	0.14	0.10	0.00	-0.01	0.10	-0.05	-0.02	-0.08	0.12	-0.05	0.02	0.10	-0.04	-0.03
20	-0.06	0.00	0.02	0.08	-0.09	-0.05	0.22	--	0.29	-0.22	0.20	0.32	0.16	0.23	0.35	0.27	0.21	0.02	0.08	-0.06	0.11	0.10	0.21	0.12	0.01	0.18
22	0.04	0.00	0.18	0.05	-0.09	0.40	0.13	0.20	--	0.12	0.69	0.24	0.00	0.08	0.37	0.71	0.72	0.52	0.48	0.57	-0.02	0.37	0.29	0.19	0.06	0.17
23	0.18	0.46	0.17	0.26	0.38	0.12	0.44	-0.02	0.07	--	0.13	-0.17	0.00	0.02	0.10	0.13	0.09	0.35	0.18	0.07	-0.07	0.18	-0.08	0.07	0.10	-0.14
26	0.00	0.03	0.09	0.02	-0.06	0.28	0.07	0.20	0.60	0.13	--	0.16	0.08	-0.05	0.35	0.80	0.66	0.51	0.56	0.42	0.10	0.39	0.25	0.04	-0.02	0.25
35	-0.09	-0.16	-0.01	-0.03	-0.13	0.15	-0.06	0.26	0.19	-0.18	0.16	--	0.22	0.10	0.30	0.19	0.17	0.11	-0.04	0.17	0.04	0.03	0.08	0.06	0.24	0.22
39	-0.05	-0.04	0.24	0.15	0.02	0.08	0.07	0.13	0.03	-0.04	0.01	0.24	--	0.07	0.16	-0.14	0.00	0.16	-0.02	-0.01	0.07	-0.06	0.02	0.01	0.25	0.18
43	0.16	0.06	0.07	0.10	0.11	0.23	0.06	0.07	0.19	0.15	0.05	0.20	-0.01	--	-0.03	0.02	0.18	0.02	-0.04	0.05	-0.06	0.07	0.47	-0.10	0.33	-0.11
45	-0.13	0.02	0.14	0.04	-0.12	0.17	0.00	0.27	0.29	-0.01	0.33	0.33	0.24	-0.01	--	0.26	0.20	0.07	0.03	0.02	-0.07	0.10	0.01	-0.03	0.04	0.04
38	0.01	-0.03	0.01	0.02	-0.07	0.30	0.05	0.18	0.59	0.04	0.70	0.05	-0.16	0.02	0.15	--	0.79	0.69	0.80	0.65	0.09	0.38	0.40	0.10	0.04	0.29
11	0.00	-0.02	-0.01	-0.01	-0.16	0.33	0.15	0.14	0.67	0.09	0.63	0.14	-0.04	0.15	0.24	0.71	--	0.66	0.61	0.65	0.00	0.46	0.37	0.03	0.21	0.25
44	0.11	0.07	0.21	0.14	0.04	0.47	0.13	0.16	0.62	0.14	0.51	0.00	0.02	0.16	0.13	0.70	0.66	--	0.63	0.77	-0.08	0.69	0.42	-0.07	0.36	0.25
17	0.02	-0.03	-0.02	0.01	-0.06	0.25	-0.02	0.12	0.46	0.06	0.49	-0.01	-0.19	-0.01	0.10	0.79	0.58	0.70	--	0.67	-0.05	0.45	0.33	0.04	0.02	0.11
47	0.02	-0.02	0.14	0.01	-0.11	0.40	0.08	0.11	0.62	0.01	0.45	0.06	-0.01	0.10	0.09	0.64	0.63	0.71	0.63	--	-0.12	0.33	0.35	0.28	0.20	0.24
40	0.06	-0.03	0.05	-0.03	0.01	0.06	0.14	0.15	0.04	-0.07	0.06	-0.03	-0.05	-0.09	0.06	0.08	0.05	0.11	0.08	0.02	--	-0.09	0.04	0.52	-0.10	0.02
15	-0.01	0.05	0.06	0.05	-0.09	0.37	0.04	0.11	0.45	0.13	0.38	0.03	-0.03	0.16	0.21	0.47	0.50	0.61	0.47	0.49	0.03	--	0.30	0.12	0.13	0.00
34	0.07	0.05	-0.05	0.04	-0.07	0.34	0.02	0.17	0.37	-0.07	0.27	0.10	-0.04	0.30	0.06	0.30	0.34	0.47	0.31	0.44	0.09	0.32	--	0.17	0.46	0.20
50	-0.01	0.02	0.01	-0.04	-0.06	0.02	0.04	0.15	0.21	-0.02	0.05	0.04	-0.06	0.03	0.06	0.12	0.23	0.12	0.17	0.15	0.44	0.22	0.10	--	0.14	0.02
25	0.10	-0.01	0.02	0.01	0.05	0.19	-0.02	0.03	0.19	0.09	0.06	0.21	0.14	0.42	0.04	0.06	0.18	0.21	0.08	0.23	0.00	0.19	0.40	0.09	--	0.19
37	-0.09	-0.10	-0.10	-0.14	-0.10	-0.07	-0.04	0.21	0.10	-0.09	0.16	0.19	0.12	-0.08	0.11	0.17	0.20	0.04	0.11	0.13	0.05	0.02	0.17	0.05	0.14	--

Note. Bottom half of the table contains polychoric correlations (4-point). Top half contains tetrachoric correlations (2-point). 74% of polychorics are below |0.2| and 67% of tetrachorics are below |0.2|

Table 3

Response Percentages for the AQ-26

	Item	4-point				2-point	
		1	2	3	4	1	2
5	I often notice small sounds when others do not.	11.1	29.5	**40.1**	19.3	40.6	**59.4**
6	I usually notice car number plates or similar strings of information.	21.7	28.4	**34.3**	15.6	**50.1**	49.9
7	Other people frequently tell me that what I've said is impolite, even though I think it is polite.	**65.1**	24.3	9.5	1.1	**89.4**	10.6
9	I am fascinated by dates.	**42.7**	36.0	16.7	4.6	**78.7**	21.3
11[a]	I find social situations easy.	3.9	15.4	**47.3**	33.4	19.3	**80.7**
12	I tend to notice details that others do not.	2.2	16.1	**51.0**	30.8	18.3	**81.8**
13	I would rather go to a library than a party.	**46.0**	31.0	17.8	5.2	**77.0**	23.0
15[a]	I find myself drawn more strongly to people than to things.	0.9	16.3	**44.5**	38.4	17.2	**82.9**
17	I enjoy social chit-chat.	2.8	11.5	39.7	**46.0**	14.3	**85.7**
19	I am fascinated by numbers.	**38.0**	28.9	26.7	6.5	**66.9**	33.2
20	When I'm reading a story, I find it difficult to work out the characters' intentions.	29.1	**54.4**	14.8	1.7	**83.5**	16.5
22	I find it hard to make new friends.	**53.6**	28.9	13.2	4.3	**82.5**	17.5
23	I notice patterns in things all the time.	6.7	29.3	**46.4**	17.6	36.0	**64.0**
25[a]	It does not upset me if my daily routine is disturbed.	14.8	**38.6**	**34.5**	12.1	**53.4**	46.6
26	I frequently find that I don't know how to keep a conversation going.	37.3	**40.8**	17.4	4.6	**78.1**	22.0
34[a]	I enjoy doing things spontaneously.	2.6	13.9	**44.0**	39.5	16.5	**83.5**
35	I am often the last to understand the point of a joke.	24.7	**42.7**	22.3	10.2	**67.4**	32.5
37[a]	If there is an interruption, I can switch back to what I was doing very quickly.	5.9	37.3	**46.0**	10.8	43.2	**56.8**
38[a]	I am good at social chit-chat.	3.7	13.4	**43.2**	39.7	17.1	**82.9**
39	People often tell me that I keep going on and on about the same thing.	26.7	**45.1**	22.8	5.4	**71.8**	28.2
40[a]	When I was young, I used to enjoy playing games involving pretending with other children.	6.1	9.5	31.9	**52.5**	15.6	**84.4**
43	I like to plan any activities I participate in carefully.	4.1	27.1	**46.2**	22.6	31.2	**68.8**
44[a]	I enjoy social occasions.	0.4	3.3	26.9	**69.4**	3.7	**96.3**
45	I find it difficult to work out people's intentions.	17.1	**52.9**	25.4	4.6	**70.0**	30.0
47[a]	I enjoy meeting new people.	0.2	6.1	33.2	**60.5**	6.3	**93.7**
50[a]	I find it very easy to play games with children that involve pretending.	5.9	16.3	37.7	**40.1**	22.2	**77.8**

Response percentages correspond to the original items, PRIOR to reverse coding negative items.

Scale: 1 - Strongly Disagree 2 - Slightly Disagree 3 - Slightly Agree 4 - Definitely Agree. The item numbering is not consecutive and corresponds to the original numbering of items in AQ-50.

[a] - reverse scored

Table 4

CFA Fit Indices for the AQ-26 3-F Model (N = 461)

Model	4-point							2-point						
	χ^2	*df**	Free Parameters	CFI	TLI	RMSEA	WRMR	χ^2	*df**	Free Parameters	CFI	TLI	RMSEA	WRMR
Single-Factor[1]	1521.279**	106	103	0.501	0.619	0.170	2.892	705.192**	117	51	0.468	0.540	0.104	2.100
Five-Factor[2]	675.302**	106	114	0.799	0.847	0.108	1.855			Inadmissable solution				
Three-Factor[3]	499.185**	102	107	0.860	0.889	0.092	1.685	266.919**	110	55	0.858	0.870	0.056	1.321
Higher-order[4]			Inadmissable solution							Inadmissable solution				
Four-Factor[5]	473.873**	102	110	0.869	0.896	0.089	1.620			Inadmissable solution				

Note. *Degrees of freedom are calculated according to the formula 110 (p. 358) in the Mplus User's Guide (Muthén & Muthén, 1998a and Muthén & Muthén, 1998b).

**p < .001

CFI = comparative fit index; TLI = Tucker Lewis index; RMSEA = root mean square error of approximation; WRMR = weighted root mean square residual

1 - uni-dimensional model under which a total AQ-26 score is calculated.

2 - model theorized by Baron-Cohen et al. (2001).

3 - model championed by Austin (2005) and partially supported by Hurst et al. (2007).

4 - model championed by Hoekstra et al. (2008).

5 - model championed by Stewart & Austin (2009).

Table 5

Correlation Residuals for Misspecified Item Relationships

Item (Factor)		Correlation Residuals		Possible reason
		4-point	2-point	
25[a] It does not upset me if my daily routine is disturbed. 34[a] I enjoy doing things spontaneously.	(DP) (SS)	0.387	0.458	Both items pertain to routine and spontaneity.
25[a] It does not upset me if my daily routine is disturbed. 43 I like to plan any activities I participate in carefully .	(DP) (DP)	0.358	0.327	Both items pertain to routine and planning.
40[a] When I was young I used to enjoy playing games involving pretending with other children. 50[a] I find it very easy to play games with children that involve pretending.	(SS) (SS)	0.411	0.516	Both items pertain to playing pretend games with children.
25[a] It does not upset me if my daily routine is disturbed. 44[a] I enjoy social occasions.	(DP) (SS)		0.358	No identifiable reason.
34[a] I enjoy doing things spontaneously. 43 I like to plan any activities I participate in carefully.	(SS) (DP)		0.468	Both items pertain to spontaneity and planning.
7. Other people frequently tell me that's what I said is impolite although I think it's polite. 22. I find it hard to make new friends.	(SS) (CM)		0.328	No identifiable reason.

Note. Positive residuals indicate underestimated relationships. [a] - reverse scored item.
SS = Social Skills; DP = Details/Patterns; CM = Communication/Mindreading

Table 6

EFA Fit Indices for the AQ-26 (N = 461)

Factor	χ^2	*df*	Free parameters	CFI	TLI	RMSEA	SRMR
1	925.674	111	26	0.713	0.791	0.126	0.108
2	598.866	117	51	0.831	0.882	0.095	0.078
3	446.279	115	75	0.883	0.918	0.079	0.062
4	331.299	114	98	0.923	0.946	0.064	0.049
5	258.468	111	120	0.948	0.962	0.054	0.041
6	216.097	104	141	0.960	0.969	0.048	0.036
7	189.132	98	161	0.968	0.973	0.045	0.032
8	153.644	90	180	0.978	0.980	0.039	0.027

Table 7

Pattern Coefficients for the 3-, 4-, 5- and 6-Factor EFA Models (N = 461)

		3-factor			4-factor				5-factor					6-factor					
		2	1	3	2	1	3	5	2	1	3	5	6	2	1	3	4	5	6
23	I notice patterns in things all the time.	0.04	**0.67**	-0.01	0.05	**0.67**	-0.03	-0.02	0.09	**0.67**	-0.03	-0.01	-0.04	0.07	**0.73**	-0.02	-0.09	0.05	-0.06
6	I usually notice car number plates or similar strings of information.	-0.04	**0.66**	0.02	-0.02	**0.67**	-0.01	-0.02	-0.01	**0.67**	-0.03	0.00	0.04	-0.01	**0.63**	-0.08	0.08	0.00	0.03
19	I am fascinated by numbers.	0.03	**0.54**	0.13	0.03	**0.60**	0.23	-0.12	0.00	**0.61**	0.20	-0.09	0.20	-0.01	**0.62**	0.16	0.05	-0.06	0.20
12	I tend to notice details that others do not.	-0.14	**0.52**	-0.02	-0.12	**0.50**	-0.14	0.08	-0.10	**0.50**	-0.14	0.08	-0.04	-0.10	**0.48**	-0.17	0.00	0.10	-0.03
9	I am fascinated by dates.	-0.04	**0.50**	0.18	-0.04	**0.53**	0.16	0.02	-0.02	**0.53**	0.15	0.02	0.00	-0.02	**0.45**	0.05	0.30	-0.04	0.02
5	I often notice small sounds when others do not.	0.00	**0.38**	-0.01	0.01	**0.35**	-0.17	0.16	0.00	**0.35**	-0.18	0.17	0.02	0.02	**0.32**	-0.21	0.03	0.16	0.03
38	I am good at social chit-chat.	**0.96**	-0.03	-0.28	**0.94**	-0.03	0.00	-0.21	**0.95**	-0.03	0.01	-0.21	-0.03	**0.94**	0.01	0.09	-0.13	-0.19	-0.04
17	I enjoy social chit-chat.	**0.88**	0.01	**-0.35**	**0.87**	-0.04	-0.18	-0.11	**0.87**	-0.04	-0.18	-0.11	0.02	**0.87**	-0.06	-0.11	-0.04	-0.14	0.02
44	I enjoy social occasions.	**0.85**	0.22	-0.02	**0.82**	0.16	-0.06	0.19	**0.81**	0.16	-0.06	0.20	0.02	**0.85**	0.02	-0.10	**0.33**	0.03	0.04
47	I enjoy meeting new people.	**0.78**	0.03	0.00	**0.75**	-0.03	-0.02	0.19	**0.75**	-0.03	-0.02	0.19	-0.01	**0.76**	-0.12	-0.02	0.22	0.06	-0.01
11	I find social situations easy.	**0.81**	-0.03	0.05	**0.77**	-0.03	0.19	0.04	**0.76**	-0.03	0.18	0.04	0.03	**0.75**	0.00	0.23	-0.01	0.04	0.02
26	I frequently find that I don't know how to keep a conversation going.	**0.71**	-0.01	0.04	**0.67**	0.04	**0.30**	-0.14	**0.70**	0.04	**0.31**	-0.16	-0.06	**0.66**	0.10	**0.37**	-0.04	-0.12	-0.07
22	I find it hard to make new friends.	**0.70**	0.02	0.21	**0.66**	0.04	0.29	0.12	**0.65**	0.03	0.28	0.12	0.02	**0.65**	0.01	0.28	0.17	0.06	0.02
15	I find myself drawn more strongly to people than to things.	**0.60**	0.08	0.04	**0.58**	0.05	0.03	0.16	**0.56**	0.05	0.02	0.16	0.05	**0.58**	-0.03	0.01	0.20	0.07	0.05
13	I would rather go to a library than a party.	**0.37**	0.25	0.23	**0.35**	0.22	0.09	0.28	**0.36**	0.22	0.10	0.27	-0.06	**0.38**	0.09	0.00	**0.41**	0.15	-0.05
34	I enjoy doing things spontaneously.	**0.41**	0.01	0.25	**0.38**	-0.07	-0.01	**0.46**	**0.34**	-0.07	-0.02	**0.48**	0.09	**0.36**	-0.06	-0.01	0.04	**0.45**	0.10
45	I find it difficult to work out people's intentions.	0.16	-0.13	**0.38**	0.12	-0.04	**0.56**	-0.05	0.12	-0.04	**0.57**	-0.06	0.04	0.10	-0.03	**0.53**	0.19	-0.08	0.03
35	I am often the last to understand the point of a joke.	-0.01	-0.30	**0.57**	-0.05	-0.24	**0.50**	0.25	-0.05	-0.24	**0.52**	0.24	-0.01	-0.08	-0.16	**0.52**	0.04	0.27	-0.02
20	When I'm reading a story, I find it difficult to work out the characters' intentions.	0.14	-0.08	**0.30**	0.11	-0.01	**0.43**	-0.04	0.05	0.00	**0.41**	-0.02	0.27	0.02	0.10	**0.46**	-0.09	0.05	0.26
39	People often tell me that I keep going on and on about the same thing.	-0.19	-0.03	**0.47**	-0.21	0.04	**0.45**	0.12	-0.19	0.03	**0.47**	0.10	-0.05	-0.21	-0.03	**0.39**	**0.34**	0.03	-0.04
37	If there is an interruption, I can switch back to what I was doing very quickly.	0.14	-0.23	0.15	0.12	-0.21	0.22	0.02	0.09	-0.20	0.21	0.02	0.09	0.06	-0.08	**0.31**	-0.23	0.11	0.08
7	Other people frequently tell me that what I've said is impolite, even though I think it is polite.	0.00	0.28	0.20	0.00	**0.32**	0.20	0.03	0.01	**0.32**	0.20	0.02	0.00	0.03	0.19	0.08	**0.42**	-0.11	0.01
25	It does not upset me if my daily routine is disturbed.	0.10	0.05	**0.42**	0.06	-0.04	0.04	**0.65**	0.03	-0.04	0.04	**0.66**	0.00	0.04	0.03	0.06	-0.04	**0.70**	0.01
43	I like to plan any activities I participate in carefully.	0.04	0.18	**0.38**	0.01	0.12	0.00	**0.59**	0.01	0.11	0.01	**0.58**	-0.08	0.02	0.16	0.01	0.00	**0.60**	-0.08
40	When I was young, I used to enjoy playing games involving pretending with other children.	0.12	-0.01	-0.01	0.12	0.01	0.08	-0.08	-0.06	0.04	-0.01	-0.01	**0.71**	-0.04	0.01	-0.02	0.01	-0.05	**0.73**
50	I find it very easy to play games with children that involve pretending.	0.21	-0.04	0.03	0.20	-0.03	0.09	-0.01	0.05	-0.02	-0.01	0.08	**0.60**	0.08	-0.02	0.01	-0.03	0.06	**0.57**

Inter-factor Correlations

	2	1	3
2	--		
1	0.08	--	
3	0.08	0.09	--

	2	1	3	5
2	--			
1	0.06	--		
3	0.16	0.00	--	
5	0.15	0.09	0.12	--

	2	1	3	5	6
2	--				
1	0.03	--			
3	0.15	0.02	--		
5	0.19	0.08	0.12	--	
6	0.21	-0.06	0.01	0.02	--

	2	1	3	4	5	6
2	--					
1	0.06	--				
3	0.01	0.21	--			
4	0.15	-0.06	0.07	--		
5	0.21	-0.01	0.23	0.06	--	
6	0.18	-0.04	0.03	0.02	0.04	--

Note. Factor numbers correspond to the six-factor solution.

Table 8

Pattern Coefficients from the Six Factor EFA and the Three-Factor Model Supported by Austin (2005) and Hurst et al. (2007) (N = 461)

	Item	Three-factor (Austin, 2005; Hurst et al., 2007)		2	1	3	4	5	6
23	I notice patterns in things all the time.		V	0.07	**0.73**	-0.02	-0.09	0.05	-0.06
6	I usually notice car number plates or similar strings of information.		V	-0.01	**0.63**	-0.08	0.08	0.00	0.03
9	I am fascinated by numbers.		V	-0.01	**0.62**	0.16	0.05	-0.06	0.20
12	I tend to notice details that others do not.		V	-0.10	**0.48**	-0.17	0.00	0.10	-0.03
19	I am fascinated by dates.		V	-0.02	**0.45**	0.05	0.30	-0.04	0.02
5	I often notice small sounds when others do not.		V	0.02	**0.32**	-0.21	0.03	0.16	0.03
25	It does not upset me if my daily routine is disturbed.		V	0.04	0.03	0.06	-0.04	**0.70**	0.01
43	I like to plan any activities I participate in carefully.		V	0.02	0.16	0.01	0.00	**0.60**	-0.08
38	I am good at social chit-chat.	V		**0.94**	0.01	0.09	-0.13	-0.19	-0.04
17	I enjoy social chit-chat.	V		**0.87**	-0.06	-0.11	-0.04	-0.14	0.02
44	I enjoy social occasions.	V		**0.85**	0.02	-0.10	**0.33**	0.03	0.04
47	I enjoy meeting new people.	V		**0.76**	-0.12	-0.02	0.22	0.06	-0.01
11	I find social situations easy.	V		**0.75**	0.00	0.23	-0.01	0.04	0.02
26	I frequently find that I don't know how to keep a conversation going.	V		**0.66**	0.10	**0.37**	-0.04	-0.12	-0.07
22	I find it hard to make new friends.	V		**0.65**	0.01	0.28	0.17	0.06	0.02
15	I find myself drawn more strongly to people than to things.	V		**0.58**	-0.03	0.01	0.20	0.07	0.05
13	I would rather go to a library than a party.	V		**0.38**	0.09	0.00	**0.41**	0.15	-0.05
34	I enjoy doing things spontaneously.	V		**0.36**	-0.06	-0.01	0.04	**0.45**	0.10
40	When I was young, I used to enjoy playing games involving pretending with other children.	V		-0.04	0.01	-0.02	0.01	-0.05	**0.73**
50	I find it very easy to play games with children that involve pretending.	V		0.08	-0.02	0.01	-0.03	0.06	**0.57**
45	I find it difficult to work out people's intentions.		V	0.10	-0.03	**0.53**	0.19	-0.08	0.03
35	I am often the last to understand the point of a joke.		V	-0.08	-0.16	**0.52**	0.04	0.27	-0.02
20	When I'm reading a story, I find it difficult to work out the characters' intentions.		V	0.02	0.10	**0.46**	-0.09	0.05	0.26
39	People often tell me that I keep going on and on about the same thing.		V	-0.21	-0.03	**0.39**	**0.34**	0.03	-0.04
37	If there is an interruption, I can switch back to what I was doing very quickly.		V	0.06	-0.08	**0.31**	-0.23	0.11	0.08
7	Other people frequently tell me that what I've said is impolite, even though I think it is		V	0.03	0.19	0.08	**0.42**	-0.11	0.01

Figure 1

Single-Factor Model

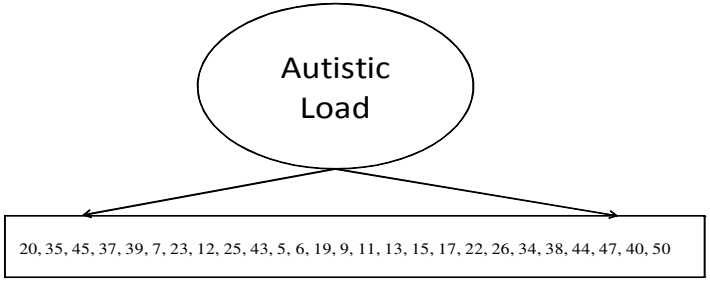


Figure 2

Five-Factor Model (Baron-Cohen et al., 2001)

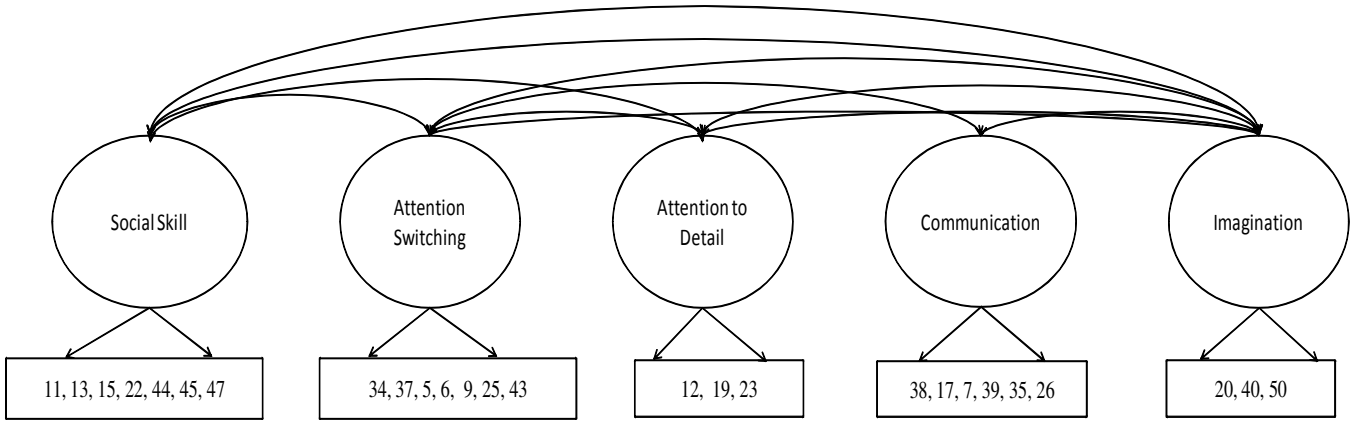


Figure 3

Three-Factor Model (Austin, 2005; Hurst et al., 2007)

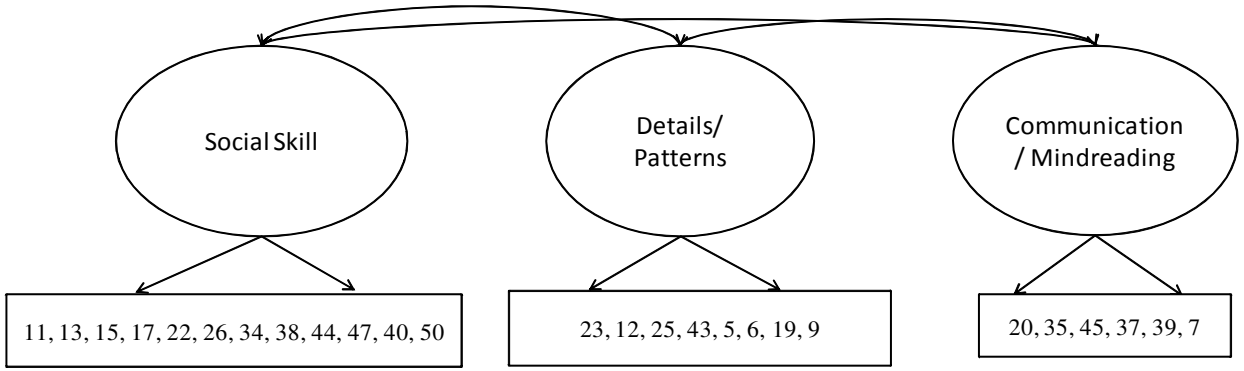


Figure 4

Higher-Order Factor Model (Hoekstra et al., 2008)

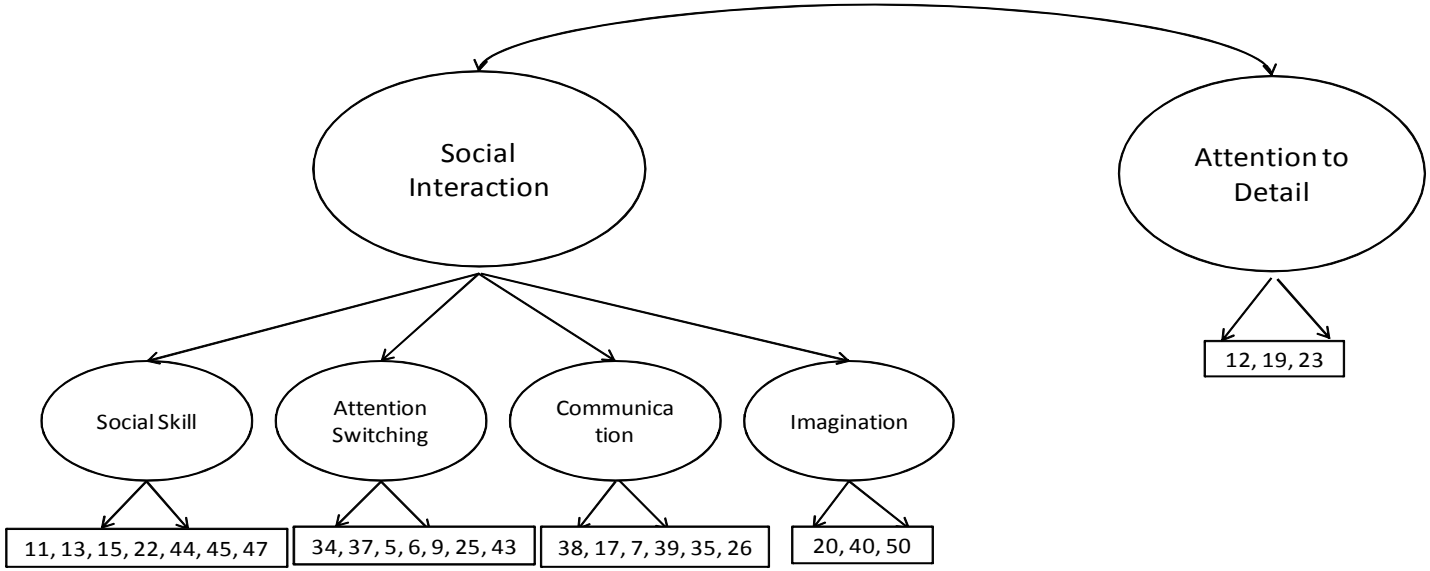
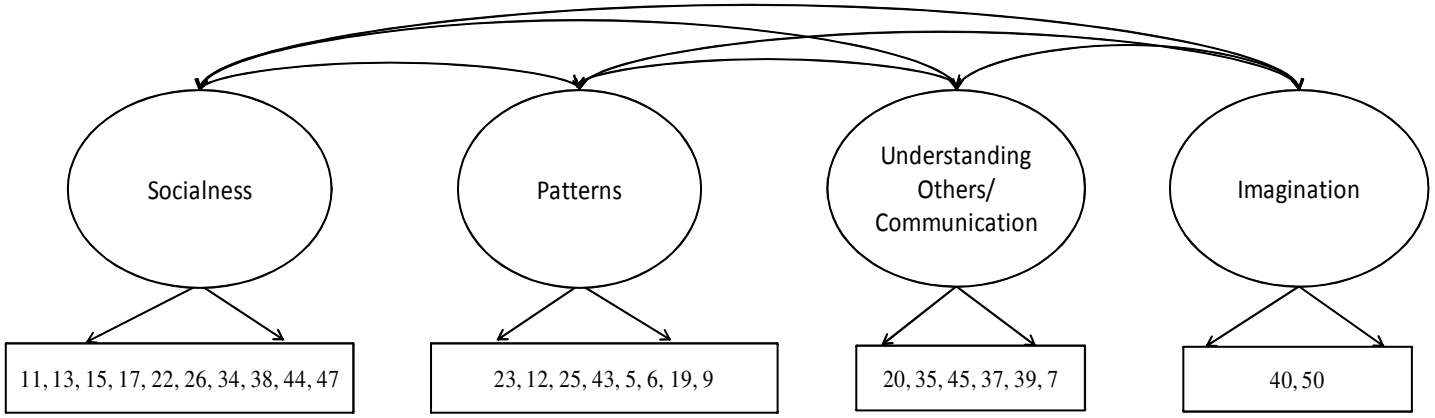


Figure 5

Four-Factor Model (Stewart & Austin, 2009)



Appendices

Appendix A

Autism-Spectrum Quotient (AQ-26)

5	I often notice small sounds when others do not.
6	I usually notice car number plates or similar strings of information.
7	Other people frequently tell me that what I've said is impolite, even though I think it is polite.
9	I am fascinated by dates.
11[a]	I find social situations easy.
12	I tend to notice details that others do not.
13	I would rather go to a library than a party.
15[a]	I find myself drawn more strongly to people than to things.
17[a]	I enjoy social chit-chat.
19	I am fascinated by numbers.
20	When I'm reading a story, I find it difficult to work out the characters' intentions.
22	I find it hard to make new friends.
23	I notice patterns in things all the time.
25[a]	It does not upset me if my daily routine is disturbed.
26	I frequently find that I don't know how to keep a conversation going.
34[a]	I enjoy doing things spontaneously.
35	I am often the last to understand the point of a joke.
37[a]	If there is an interruption, I can switch back to what I was doing very quickly.
38[a]	I am good at social chit-chat.
39	People often tell me that I keep going on and on about the same thing.
40[a]	When I was young, I used to enjoy playing games involving pretending with other children.
43	I like to plan any activities I participate in carefully.
44[a]	I enjoy social occasions.
45	I find it difficult to work out people's intentions.
47[a]	I enjoy meeting new people.
50[a]	I find it very easy to play games with children that involve pretending.

Scale: 1 - Strongly Disagree 2 - Slightly Disagree 3 - Slightly Agree 4 - Definitely Agree.
The item numbering is not consecutive and corresponds to the original numbering of items in AQ-50.
[a] - reverse scored

Appendix B

DSM-IV-*TR* Diagnostic Criteria for Asperger's Disorder

(I) Qualitative impairment in social interaction, as manifested by at least two of the following:

- (A) marked impairments in the use of multiple nonverbal behaviors such as eye-to-eye gaze, facial expression, body posture, and gestures to regulate social interaction
- (B) failure to develop peer relationships appropriate to developmental level
- (C) a lack of spontaneous seeking to share enjoyment, interest or achievements with other people, (e.g.. by a lack of showing, bringing, or pointing out objects of interest to other people)
- (D) lack of social or emotional reciprocity

(II) Restricted repetitive & stereotyped patterns of behavior, interests and activities, as manifested by at least one of the following:

- (A) encompassing preoccupation with one or more stereotyped and restricted patterns of interest that is abnormal either in intensity or focus
- (B) apparently inflexible adherence to specific, nonfunctional routines or rituals
- (C) stereotyped and repetitive motor mannerisms (e.g. hand or finger flapping or twisting, or complex whole-body movements)
- (D) persistent preoccupation with parts of objects

(III) The disturbance causes clinically significant impairments in social, occupational, or other important areas of functioning.

(IV) There is no clinically significant general delay in language (E.G. single words used by age 2 years, communicative phrases used by age 3 years)

(V) There is no clinically significant delay in cognitive development or in the development of age-appropriate self help skills, adaptive behavior (other than in social interaction) and curiosity about the environment in childhood.

(VI) Criteria are not met for another specific Pervasive Developmental Disorder or Schizophrenia.

References

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

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

Americans with Disabilities Act of 1990, Pub. L. No. 101-336, § 2, 104 Stat. 328 (1991; 2008).

Asperger, H. (1944). Die autistischen Psychopathen im Kindesalter. *Archiv fur Psychiatrie und Nervenkrankheiten, 177,* 76-137.

Attwood, T. (2007). *The complete guide to Asperger's syndrome*. London: Jessica Kingsley Publishers.

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ). *Personality and Individual Differences, 38(2)*, 451-460.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5-17.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2005). The Adult Asperger Assessment (AAA): a diagnostic method. *Journal of Autism and Developmental Disorders, 35,* 817-819.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17*(1), 10-17.

Benson, J., & Nasser, F. (1998). On the use of factor analysis as a research tool. *Journal of Vocational Education Research, 23*(1), 13-33.

Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling, 7*(3), 461-483.

Bollen, K. A. (2000). Modeling strategies: In search of the holy grail. *Structural Equation Modeling, 7*, 74 – 81.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Centers for Disease Control. (2007). *Prevalence of autism spectrum disorders – Autism and developmental disabilities monitoring network. 14 sites, United States, 2002.* (Surveillance Summaries MMWR No. SS-1). Atlanta, GA: Author.

Dunn, W., Myles, B.S., & Orr, S. (2002). Sensory processing issues associated with Asperger Syndrome: A preliminary investigation. *The American Journal of Occupational Therapy, 56 (1)*, 97–102.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis*. Cambridge, MA: The MIT Press.

Foster, M. A. (2003). *The development of the adult Asperger syndrome scale* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3094777)

Finney, S. J., & DiStefano, C. (2006). Nonnormal and categorical data in structural equation models. In G. R. Hancock & R. O. Mueller (Eds.). *A second course in structural equation modeling* (pp. 269 - 314). Greenwich, CT: Information Age.

Flora, D. B. & Curran, P. J. (2004) An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466-491.

Frith, U. (1989). *Autism: Explaining the Enigma.* Oxford: Basil Blackwell Ltd.

Grossman, P. D. (2001). Making accommodations: The legal world of students with disabilities.*87 Academe: Bulletin of the American Association of University Professors,* 41-46.

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

Gillberg, C. & Coleman, M. (2000) *The Biology of the Autistic Syndromes, Third Edition.* Cambridge University Press.

Gillberg C., Gillberg C., Rastam, M., & Wentz, E. (2001). The Asperger Syndrome (and high-functioning autism) Diagnostic Interview (ASDI): a preliminary study of a new structured clinical interview. *Autism* , 5, 57–66.

Glennon, T. J. (2001). The stress of the university experience for students with Asperger syndrome. *Journal of Prevention, Assessment & Rehabilitation, 17*(3),189-190.

Joreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202.

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validit of the Autism-Spectrum Quotient (AQ): A study in

Dutch population and patient groups. *Journal of Autism and Developmental Disorders, 38*, 1555-1566.

Howlin, P. (2000). Assessment Instruments for Asperger Syndrome. *Child Psychology and Psychiatry Review, 5*(3), 120-129.

Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: Implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders, 33* (1), 3–13.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424-453.

Hurst, R. M., Mitchell, J. T., Kimbrel, N. A., Kwapil, T. K., & Nelson-Gray, R. O. (2007). Examination of the reliability and factor structure of the Autism Spectrum Quotient (AQ) in a non-clinical sample. *Personality and Individual Differences, 43*, 1938-1949.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2,* 217-250.

Kline, R. B. (2005). *Structural Equation Modeling*. New York, NY: Guilford Press.

Liss, M., Mailloux, J., & Erchull, M. J. (2008). The relationships between sensory processing sensitivity, alexithymia, autism, depression, and anxiety. *Personality and Individual Differences, 45*(3), 255-259.

Matson, J. L., & LoVullo, S. V. (2009). Trends and topics in autism spectrum disorders research. *Research in Autism Spectrum Disorders, 3*, 252-257.

Matson, J. L. & Boijoli J. A. (2008). Strategies for assessing Asperger's syndrome: A critical review of data-based methods. *Research in Autism Spectrum Disorders, 2,* 237-248.

Myles, B., Bock, S., & Simpson, S. (2001). Asperger Syndrome Diagnostic Scale. In Pro-ed (Ed.). Austin, TX.

Myles, B. S., & Simpson, R. L. (2002). Asperger Syndrome: An overview of characteristics. *Focus on Autism and Other Developmental Disorders, 17*(3).

Myles, B. S., Cook, K. T., Miller, N. E., Rinner, L., & Robbins, L. A. (2000). *Asperger syndrome and sensory issues: Practical solutions for making sense of the world.* Shawnee Mission, KS: AAPC.

Muthen, B. O. (1993). Goodness of fit with categorical and other nonnormal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205-234). Newbury Park, CA: Sage.

Muthen, B. O., du Toit, S. H. C., Spisic, D. (in press). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika.*

Muthen, L. K., & Muthen, B. O. (2001). *Mplus user's guide.* Los Angeles: Author.

Ollson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrica, 44*, 443-460.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2,* 13-32.

Riggio, R. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology*, 51(3), 649-660.

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 Asperger's disorder in adults (RAADS): A pilot study. *Journal of Autism and Developmental Disorders, 38*, 213-223.

Schutte, J., Angelone, D. A., McCue M., & McGonigle, J. (2008). *Rehabilitation and Vocational Counseling for Adults with Asperger's Syndrome.* Unpublished manuscript, University of Pittsburgh, PA, USA.

Smith, I. (2000). Motor functioning in Asperger syndrome. In A. Klin, F. Volkmar, & S. Sparrow (Eds.), *Asperger Syndrome.* New York: the Guilford Press.

Smith, C. P. (2007). Support services for students with Asperger's Syndrome in higher education. *College Student Journal, 41*(3), 515-531.

Steenkamp, J., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78 – 90.

Stewart, M. E., & Austin, E. J. (2009). The structure of the Autism-Spectrum Quotient (AQ): Evidence from a student sample in Scotland. *Personality and Individual Differences, 47,* 224-228.

Tantam, D. (2000). Adolescence and adulthood of individuals with Asperger syndrome. In A. Klin, F. R. Volkmar, & S. S. Sparrow (Eds.), *Asperger syndrome* (pp. 367-399). New York: The Guildford Press.

Voracek, M., & Dressler, S. G. (2006). Lack of correlation between digit ratio (2D : 4D) and Baron-Cohen's "Reading the Mind in the Eyes" test, empathy, systemising, and autism-spectrum quotients in a general population sample. *Personality and Individual Differences, 41*(8), 1481-1491.

West, S.G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications* (p. 57 – 75). Thousand Oaks, CA: Sage.

Wing, L. (1981). Asperger's Syndrome: a clinical account. *Psychological Medicine, 11,* 115-130.

World Health Organization (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research.* Geneva: WHO.

Yu, C., & Muthén, B (2002, April). *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Yuan, K.-H., Marshall, L.L., & Bentler, P.M. (2003). Assessing the effect of model misspecifications on parameter estimates in structural equation models. Sociological Methodology, 33, 241–265.