The Factor Structure of the Broad Autism Phenotype Questionnaire in a Non-Clinical Sample of Canadian Undergraduate Students

by

Sydney E. Kingston

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

University of Manitoba Winnipeg

Copyright © 2024 by Sydney Kingston

Abstract

Over the last several decades there has been increasing recognition that symptoms of Autism Spectrum Disorder (ASD) vary along a continuum from absent to present and mild to severe. Relatives of those with autism may present with subclinical traits that do not meet diagnostic thresholds, but may nevertheless reflect challenges like those of people diagnosed with ASD. The Broad Autism Phenotype (BAP) reflects a dimensional approach to ASD that includes characteristics of the 'milder but qualitatively similar' (Hurley et al, 2007) presentation that is often seen in family members of those with ASD (Happé & Frith, 2020; Robinson, Koenen et al., 2011) as well as the general population (Constantino & Todd, 2003). Understanding the characteristics of this subclinical group may be important for research and clinical contexts, including identifying affected individuals and selecting interventions that may be helpful for functional impairments. As such, accurate measurement of the BAP is critical. Various questionnaires have been used for assessing the BAP; however, the Broad Autism Phenotype Questionnaire (BAPQ; Hurley et al., 2007) is one of the most widely used (Beaver, 2018). Despite its wide use, the limited existing psychometric research on the BAPQ is mixed. Some research suggests problems with its internal structure and construct validity, particularly with its Pragmatic Language subscale (e.g., Godoy-Gimenez et al., 2018; Sasson et al., 2013a). This study examines the factor structure and construct validity of the BAPQ in a sample of 899 Canadian undergraduate students who participated for course credit. Participants completed a series of online measures assessing the BAP, pragmatic language difficulties, executive functioning, and personality. Confirmatory factor analytic methods were used to examine its internal structure. Validity evidences based on relationships between BAPQ scores and other variables were examined. Four out of five fit indices examined suggested that the 3-factor model

had unacceptable fit, with Pragmatic Language items contributing to misfit. Revision and/or deletion of certain Pragmatic Language items may improve model fit. Results of this study support the convergent validity of the BAPQ. The findings of this research add to the existing literature and may have implications for the use of the BAPQ.

Keywords: Broad Autism Phenotype Questionnaire (BAPQ), psychometrics, internal structure, validity, factor analysis

Acknowledgements

I would like to thank my advisor, Dr. Janine Newton Montgomery, for her invaluable guidance, assistance, and support in completing this dissertation. I would also like to thank my advisory committee for all the thoughtful feedback they have provided throughout my journey in completing this project. I would also like to acknowledge my family, friends, and clinical psychology cohort who have provided me with unyielding support and motivation throughout the process of completing this project. Lastly, I would like to acknowledge the financial assistance that made this dissertation possible, including the University of Manitoba Graduate Fellowship and the J.G. Fletcher Award.

Table of Contents

List of Tables

List of Figures

The Factor Structure of the Broad Autism Phenotype Questionnaire in a Non-Clinical Sample of Canadian Undergraduate Students

Over the last several decades, the conceptualization and diagnosis of autism have undergone numerous changes; most notably from a once discrete diagnosis to one that is considered dimensional (Happé & Frith, 2020). Indeed, it is now recognized that autistic traits lie on a continuum from no symptoms to a clinical diagnosis of Autism Spectrum Disorder (ASD). ASD is a life-long neurodevelopmental disorder, which typically emerges during early childhood, and is characterized by a triad of main symptoms including social impairments, communication difficulties, and restricted/repetitive behaviours, interests, and activities (American Psychiatric Association [APA], 2013; Baron-Cohen & Wheelwright, 1999). Males are diagnosed with ASD more often than females at a rate of approximately 4:1, respectively (Mandy et al., 2012). Decades of research suggests that ASD is a heterogeneous disorder, with symptoms varying in presentation, type, and severity (APA, 2013). Recently, research has identified that subclinical ASD traits commonly present in family members of ASD probands can also present in the general population (e.g., Constantino & Todd, 2003; Godoy-Gimenez et al., 2018; Sasson et al., 2013a; Stewart & Austin, 2009; Wainer et al., 2011). This subclinical presentation has been termed the Broad Autism Phenotype (BAP). Indeed, the concept of the BAP suggests that there is a range of autistic traits that present not only in families of those with ASD but also in the general population. Proponents of the BAP suggest that this provides an opportunity to research ASD symptoms across ASD and Non-ASD populations. Using the BAP as another approach to ASD research may hold the potential to advance such research, particularly as it pertains to intervention. Accordingly, it is important that psychometrically sound instruments are available to assess the BAP. Various measures have been used to assess

the BAP, including the Autism Quotient (AQ; Baron-Cohen et al., 2001), the Social Responsiveness Questionnaire (SRS; Constantino et al., 2003), and the Broad Autism Phenotype Questionnaire (BAPQ; Hurley et al., 2007). Unlike the AQ and the SRS, the BAPQ was specifically developed to measure the BAP and is considered to be one of the leading questionnaires available for this purpose (Beaver, 2018). Moreover, the creators of the BAPQ suggest that it has acceptable psychometric properties (Hurley et al., 2007), and unlike the AQ (Austin, 2005; Bishop et al., 2004; Hoekstra et al., 2008; Stewart & Austin, 2009) and the SRS (Constantino et al., 2004), it has been suggested that its factor structure is consistent with its proposed theoretical model, which is a clear strength of this measure (Wainer et al., 2011). For these reasons, the BAPQ is widely used in research and clinical practice to assess BAP features. Despite its wide use, psychometric research on the BAPQ (Godoy-Gimenez et al., 2018; Hurley et al., 2007; Sasson et al., 2013a; Wainer et al., 2011) is mixed and consensus on a conceptualization of the BAP is lacking (Davidson et al., 2014; Wainer et al., 2011), with some research suggesting that its construct validity is problematic (see section on psychometric properties of the BAPQ; e.g., Godoy-Gimenez et al., 2018; Kim & Kim, 2022; Lorenz & Algner, 2021; Sasson et al., 2013a). Clearly, this has the potential to jeopardize the accuracy and relevance of research using the BAPQ as well as clinical impressions and recommendations drawn using the BAPQ in clinical practice. As such, the present study aimed to examine the internal structure and construct validity of the BAPQ in a nonclinical sample of university students attending a Canadian university to provide some additional evidence for construct validity considerations.

The Dimensional Nature of Autism and the BAP

Autism was initially conceptualized as a discrete condition and research was focussed on

identifying the pathognomonic characteristics associated with ASD to assist clinicians in making a categorical diagnosis (i.e., autism or not autism; Happé & Frith, 2020). In contrast to early work, research in recent decades suggests that ASD symptomatology is continuous (Robinson, Koenen et al., 2011; Robinson, Munir et al., 2011). Specifically, ASD is characterized by two main symptom domains: persistent impairments in social communication and interaction (e.g., difficulties maintaining a reciprocal conversation, abnormal eye contact) and restricted repetitive behaviours, interests, and activities (e.g., insistence on sameness, strict adherence to routines). As these behavioural traits are continuous, they vary in presentation, type, and severity, and the phenotypic presentation is unique for each person with ASD (APA, 2013, 2022). Supporting this spectrum, behavioural (Colvert et al., 2015; Robinson, Koenen et al., 2011) and molecular genetic (Massrali et al., 2018) research suggest that there is considerable genetic overlap between diagnosed ASD and subclinical ASD traits; with ASD representing an extreme at one end of a neurodevelopmental continuum (Robinson, Koenin et al., 2011). The genetic etiology of ASD is highly complex and no clear genetic profile has been identified as causing ASD. Hundreds of genes have been proposed to be implicated in the etiology of autism (Rylaarsdam & Guemez-Gamboa, 2019), and in many cases, ASD results from numerous genetic variants exerting small effects that produce ASD traits. Such variants are called copy number variations (CNVs) which can occur de novo or be inherited (Thapar & Cooper, 2013) and are implicated in approximately 10% of ASD cases (Geschwind, 2011). CNVs occur at the chromosomal level and consist of submicroscopic structural variants including duplications, deletions, translocations, and inversions (Marshall et al., 2008). Higher concentrations of certain variants tend to be associated with ASD symptomology (Pinto et al., 2010; Pizzo et al., 2019); however, it is important to note that CNVs are found across the entire ASD continuum and are therefore found in individuals

without ASD. Given that people possess many of these variants, it is not surprising that people without a clinical diagnosis of ASD present with subclinical features of ASD to varying degrees (Happé & Frith, 2020). This is especially evident in family members of ASD probands. Indeed, ASD is highly heritable and characteristics within the ASD symptomology triad commonly present in family members of those with ASD (Piven et al., 1997). Extensive genetic research has indicated that heritability estimates for autism range from 40% to 80% (Chaste & Leboyer, 2012).

Features of ASD in relatives of people with autism were recognized as early as 1943 with the work of Kanner, which documents unusual social behaviours in parents of children with ASD (PCA) and a preoccupation with "abstractions of a scientific literacy or artistic nature, and a limited genuine interest in people" (Kanner, 1943, p. 250). Similarly, Hans Asperger later suggested that ASD-like traits were commonly observed in the parents of children with 'high-functioning' ASD and purported that these traits were likely biologically passed from parent to offspring (Deisinger & Rotatori, 2015). Then in 1977, Folstein and Rutter conducted a twin study which supported the role of genetics in ASD etiology and was the first study to suggest that ASD traits extend beyond diagnosed cases of ASD (Losh et al., 2011; Piven, 2001). Since then, research has continued to suggest that subclinical ASD traits tend to aggregate in families and are often present in PCAs. For example, family history research on and direct assessments of relatives of people with ASD indicate that these individuals often exhibit a history of delayed language development (Szatmari et al., 2000) and impairments in pragmatic language abilities, reading, and verbal fluency (Pilowsky et al., 2003). Moreover, certain personality traits and social behaviours characteristic of ASD are frequently observed in relatives of individuals with ASD (e.g., Lainhart et al., 2003; Szatmari et al., 2000). For instance, PCAs tend to possess traits

that resemble ASD symptomatology such as social reticence, aloofness, untactful behaviour, few reciprocal friendships, lower quality friendships, rigidity, anxiety, and neuroticism (Losh et al., 2008; Micali et al., 2004). Furthermore, a systematic review identified that between 3% and 52% of PCAs demonstrate these subclinical traits (Rubenstein & Chawla, 2018). However, these phenotypic similarities tend to be milder in presentation than those present in individuals with ASD (Landry & Chouinard, 2016).

This subclinical presentation is considered an intermediate phenotype which consists of heritable subclinical traits of a particular clinical condition (Almasy & Blangero, 2001; Gershon & Goldin, 1986; Gottesman & Goulde, 2003) that can indicate genetic susceptibility for that particular condition (Iacono, 2018). In the case of ASD, this intermediate phenotype is referred to as the BAP (Goldberg et al., 2005; Piven et al., 1990; Szatmari et al., 2000) and can present across the lifespan. Characteristics of the BAP include rigidity, broadly defined stereotyped behaviours and communication difficulties, impaired social skills, pragmatic language difficulties, emotion recognition difficulties, and aloofness (Gerdts & Bernier, 2011; Sucksmith et al., 2011). These traits are qualitatively similar to ASD symptoms; however, unlike ASD, all traits may not present in those with the BAP. More specifically, a diagnosis of ASD requires that the diagnostic threshold be met in that the individual presents with the required minimum number of symptoms in each symptom category as per diagnostic criteria (e.g., APA, 2013, 2022); in contrast, someone with the BAP may only present with a subset of these symptoms/traits (Pickles et al., 2000; Szatmari et al., 2000). This suggests that the main symptom categories of ASD may be influenced by different causal factors (Ronald et al., 2006). Individuals who score above a particular threshold on a measure of the BAP are suggested to possess BAP traits and are considered BAP+. Moreover, the BAP tends to be more common in

males than females (Hurley et al., 2007; Seidman et al., 2011), which mirrors the ASD diagnostic ratio (Mandy et al., 2012).

Inherent in the term 'spectrum,' there is increasing recognition that the BAP is not restricted to relatives of ASD probands, but rather ASD symptoms and traits are continuously distributed amongst the general population (Constantino & Todd, 2003; Hoekstra et al., 2007; Landry & Chouinard, 2016; Stewart & Austin, 2009; Robinson, Munir et al., 2011; Wheelwright et al., 2010). Within this continuum, ASD-like traits can range from zero or few to enough traits to meet diagnostic criteria for ASD. For example, in a study examining the presence of elevated BAP characteristics among parents with and without a child with ASD, like PCAs, parents of children without ASD also demonstrated elevated BAP features. More specifically, 22% of fathers and 9% of mothers of a child without ASD were found to exhibit elevated BAP features compared to 33% of fathers and 23% of mothers of a child with ASD (Wheelwright et al., 2010). Considering that this sample only consisted of parents, these findings likely underestimate levels of the BAP in the general population. This underestimation is supported by research suggesting that mate selection and opportunities to reproduce may be negatively influenced by BAP traits, especially those relating to social communication and the ability to form and maintain mutually satisfying interpersonal relationships. In turn, this may result in fewer people with BAP traits reproducing and in turn becoming parents (Bailey et al., 1998). As such, studies only using samples consisting of parents may not identify more severe phenotypes and including people who are not parents may yield higher BAP estimates (Landry & Chouinard, 2016). Moreover, like ASD, BAP traits in the general population are relatively stable across time (Robinson, Munir et al., 2011).

Advantages of Studying the BAP

Individuals with ASD experience impairments across the lifespan in various domains such as social (APA, 2013, 2022), educational (e.g., Gurbuz et al., 2019), and occupational functioning (Robertson, 2010). Additionally, mental (e.g., anxiety, depression; Bellini, 2004; Hudepohl et al., 2013; Vickerstaff et al., 2007) and physical health (e.g., respiratory conditions, allergies; Gurney et al., 2006) conditions are highly prevalent in the ASD population when compared to those without ASD. Indeed, individuals with ASD tend to be higher health service users compared to those without ASD (Gurney, 2006). Taken together, these factors can adversely impact quality of life and restrict life satisfaction for these individuals. As such, research aimed at elucidating this highly complex condition and developing and enhancing existing approaches to supporting people with ASD is paramount.

Unfortunately, ASD research faces several barriers. First, it is well-accepted that ASD has a highly complex genetic etiology involving a non-Mendelian inheritance pattern in which numerous genes interact to create a predisposition to ASD (e.g., Bailey et al., 1995; Folstein & Rutter, 1977). As a consequence of this complexity and heterogeneity, traditional approaches to genetic research (e.g., genetic association research with participants with ASD) may be more difficult to conduct and potentially less fruitful. This complicates genetic research, and, consequently, the genetic etiology of ASD remains obscure (Betancur, 2011; Jeste & Geschwind, 2014). One approach to overcoming this barrier is to adopt an alternative approach to research in which intermediate phenotypes are the focus of study. This approach has "the potential to assist in the genetic dissection of psychiatric diseases" and has been used to study other genetically complex conditions such as schizophrenia (Gottesman & Gould, 2003, p. 638; Risch et al., 1999; Jones & Szatmari, 2002; Pickles et al., 1995). Given that the BAP represents an intermediate phenotype, it may offer a promising alternative approach to the genetic and biological dissection

of ASD. Indeed, examinations of the qualitative and quantitative differences (e.g., regarding genetic factors, patterns of skill and impairment) between those diagnosed with ASD and those who demonstrate the BAP may suggest factors that are uniquely involved in causing ASD (Landry & Chouinard, 2016). This is made possible by the fact that individuals with the BAP may only present with one BAP trait (e.g., rigidity) unlike ASD where multiple symptoms are required in each domain (APA, 2013, 2022). This, combined with the finding that BAP symptoms are correlated with qualitatively similar ASD symptoms, but not with dissimilar symptoms, suggests that a BAP approach may allow each symptom domain to be studied in greater isolation with greater ease and precision. Furthermore, this approach would allow researchers to target discrete areas of impairment, and, thus, findings of this research may extend to other clinical conditions with similar impairments or individuals experiencing similar impairments in the absence of a clinical diagnosis. This is in line with biological research suggesting that psychological disorders exist on a spectrum and overlap. Likewise, this approach is consistent with recommendations made by the National Institute of Mental Health (2013) and others (e.g., Adam, 2013; Craddock & Owen, 2010) to focus research on biomarkers, symptoms, and impairments instead of diagnostic categories which are consistently unsupported by genetic and neuroscientific research and are subject to change with new editions of diagnostic manuals (Adam, 2013).

Another advantage to studying the BAP is that it offers a larger pool of potential research participants compared to research using participants with ASD. Indeed, people with ASD are considered a “hard to reach” population in that recruiting participants with ASD is often time-consuming, laborious, and often results in small sample sizes compared to research with typically developing participants (Haas et al., 2016). This frequently results in smaller than

desired sample sizes as large sample sizes are often required for various statistical analyses, and, consequently, is often cited as a limitation in the ASD literature. This places limitations on conclusions that can be drawn from the results of ASD research, particularly in genetic research which requires extensive samples (Landry & Chouinard, 2016). One potential approach to addressing the sample size limitation plaguing ASD research is to recruit members from the general population who demonstrate the BAP due to large sample sizes being easier to recruit. This could significantly increase the power of these studies to identify candidate genes involved in ASD (Ingersoll & Wainer, 2014). Given the value of using the BAP as an alternative approach to ASD research and to enhancing our understanding of typical development, there is increasing research focused on BAP measurement, albeit further psychometric research is needed.

The Broad Autism Phenotype Questionnaire (BAPQ)

Various measures have been developed to quantify the BAP. Initially, the BAP was assessed via clinical interviews that were intended for use with relatives of individuals with ASD (e.g., Dawson et al., 2007). This approach soon fell out of favour in research given that clinical interviews are time intensive and require clinical training, which can be costly. Evidently, this was not a feasible measurement approach for research purposes. Within the last 20 years, various advances have been made in BAP measurement, particularly with the advent of relatively brief questionnaires designed to measure the BAP. This approach is now routinely used as it allows for the measurement of BAP traits that is time and cost-effective (Landry & Chouinard, 2016). Self-reports, informant-reports, and clinician rating scales have been developed for measuring BAP traits. Given the availability of various report forms, a multiple informant or ‘best-estimate’ approach, which collects information from multiple sources, is considered the most appropriate assessment approach (Davidson et al., 2014; Seidman et al., 2011; Sucksmith et al., 2011). Some

of the most widely used BAP questionnaires include the AQ (Baron-Cohen et al., 2001), the SRS (Constantino, 2002), and the BAPQ (Hurley et al., 2007). As previously mentioned, the BAPQ is considered one of the leading available measures of the BAP as it was specifically developed to measure the BAP (Beaver, 2018).

The BAPQ was developed using direct clinical interviews of PCAs using the Modified Personality Assessment Schedule – Revised (MPAS-R; Piven et al., 1997) and the Pragmatic Rating Scale (PRS; Hurley et al., 2007). The BAPQ consists of three subscales: Aloof, Rigid, and Pragmatic Language. These subscales reflect the ASD symptom triad detailed in the DSM-IV-TR (i.e., social impairment, language/communication impairment, and restricted/repetitive behaviours), with Aloof corresponding to social impairment, Pragmatic Language corresponding to language/communication impairment, and Rigid corresponding to restricted/repetitive behaviours. On the BAPQ, Aloof traits are defined as "a lack of interest in or enjoyment of social interaction", Pragmatic Language difficulties refer to "deficits in the social aspects of language, resulting in difficulties communicating effectively or in holding a fluid, reciprocal conversation", and Rigid traits are defined as "little interest in change or difficulty adjusting to change" (Hurley et al., 2007, p. 1681; see Methods section for additional information about the BAPQ). Considering the way that the creators of the BAPQ define Aloof traits, this subscale may only capture one of the patterns of social impairments that can be present for those on the ASD continuum. Indeed, as proposed by Wing & Gould (1979) and supported by subsequent research (e.g., Scheeren et al., 2012) three patterns of social impairment can present in ASD and tend to be associated with symptomatology severity. These patterns include: (1) "social aloofness" in which individuals tend to refrain from seeking out social interaction and may not respond to the interaction attempts of others, (2) "passive interaction" in which individuals do not seek out

social interaction but respond appropriately when socially engaged by others, and (3) “active, but odd interaction” in which individuals desire and enjoy social interaction but may engage with others in ways that are ineffective (Wing & Gould, 1979). The way in which aloof traits have been operationalized in the BAPQ appears to only reflect the first of the three social impairment patterns proposed by Wing and Gould (1979). As such, one may wonder whether the BAPQ Aloof subscale fully maps onto the full range of social impairments that may be present for those with ASD, and possibly those with the BAP.

Some research (e.g., Austin, 2005; Godoy-Gimenez et al., 2018) indicates that BAP traits overlap with other personality traits such as the five-factor model of personality (i.e., Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism; Costa & McCrae, 1992). Indeed, many early studies on the BAP focussed on the relationship between subclinical ASD traits and personality (Piven et al., 1997). Since that time, research has indicated that certain personality traits tend to be associated with BAP traits in family members of people with ASD, including shyness, aloofness, unresponsiveness, rigidity, suspiciousness, hypersensitivity to criticism, impulsivity, anxiousness, self-consciousness, and eccentricity (Lainhart et al., 2003; Losh et al., 2008; Murphy et al., 2000; Piven et al., 1997). More recently, the relationship between the BAP and the five-factor model of personality has been examined in the general population. For example, when the BAP was measured with the AQ, researchers found that BAP traits are associated with lower levels of Extraversion, Openness, Conscientiousness, and Agreeableness, and higher levels of Neuroticism (Austin, 2005; De Pauw et al., 2011; Schriber et al., 2014; Wakabayashi et al., 2006). These findings have been supported by subsequent research which assessed the relationship between the five-factor model of personality and a BAP score derived from three measures of the BAP (AQ, SRS, and BAPQ) in a college sample (Wainer et al.,

2011). Moreover, when using the BAPQ, Aloof traits (high score is interpreted as a low level of Aloof traits) have been found to be strongly and positively associated with Extraversion but moderately and positively associated with Agreeableness (Godoy-Gimenez et al., 2018). Furthermore, Rigid traits were shown to have low to moderate associations with Openness (Godoy-Gimenez et al., 2018). These research findings suggest that certain personality traits are part of the BAP's nomological network, and consequently, have appropriately been included in psychometric investigations of the BAPQ, particularly its construct validity.

Psychometric Properties of the BAPQ

Factor analysis is a method used to determine a test's internal structure and nomological network as it is developed and for subsequent psychometric analyses (e.g., construct validity evaluations). The achievement of simple structure, in which items load onto one and only one factor, is a primary aim as it indicates which items should be scored together and therefore compose a test scale reflecting a unified construct (Furr & Bacharach, 2014). Factor loadings (i.e., the association between a test item and a factor) are interpreted as correlations or regression weights and can range between -1 and +1. An examination of factor loadings helps to reveal the psychological meaning of a factor. Specifically, a factor derives its meaning based on the items that cluster together and load onto it (Furr & Bacharach, 2014). Thus, if items are assumed to measure a particular construct, they should demonstrate simple structure by only loading onto the factor that represents that construct. Items that either do not sufficiently load (i.e., have weak factor loadings) or cross-load onto other factors may not accurately capture the underlying construct, suggesting that they may need to be revised or discarded. This has implications for the scoring, evaluation, and use of a psychological measure. Despite the wide use of the BAPQ, emerging research provides mixed evidence regarding its psychometric properties. Indeed, while

some research indicates strong validity (e.g., concurrent, incremental) and reliability (e.g., internal consistency, interrater, test-retest) of the BAPQ, (e.g., Bang et al., 2022; Hurley et al., 2007; Ingersoll et al., 2011; Sasson et al., 2013) other research suggests questionable psychometric properties (Godoy-Gimenez et al., 2018; Kim & Kim, 2022; Lorenz & Algner, 2021; Nayar et al., 2018; Sasson et al., 2013a). Specific issues and psychometric challenges of the BAPQ are discussed in detail on pages 22 to 27.

Psychometric strengths. Hurley et al. (2007) report that the BAPQ has acceptable psychometric properties. More specifically, it has been validated against direct clinical interviews and assessment data and its items have been suggested to have good reliability and validity (Hurley et al., 2007; Ingersoll et al., 2011; Sasson et al., 2013). Considering that the BAPQ was specifically designed to quantify the BAP, it has convergent validity with other measures of BAP traits such as the AQ and the SRS (Hurley et al., 2007; Ingersoll et al., 2011; Sasson et al., 2013), and has acceptable internal consistency for the total scale ($\alpha = .95$) and subscales (Aloof: $\alpha = .94$; Rigid: $\alpha = .91$; Pragmatic Langauge: $\alpha = .85$). Due to being validated against clinical assessment data, it may have greater sensitivity to individual BAP traits compared to the SRS which only assesses the BAP as a single construct reflecting social functioning (Davidson et al., 2014). Furthermore, during the development of the BAPQ, optimal cut-off scores were developed for each subscale to ensure that both sensitivity and specificity were maximized. This resulted in good sensitivity and specificity; with both at or above 70% for all three subscales (Hurley et al., 2007). As the BAPQ is a continuous measure with low floors and high ceilings, it is well-suited for indexing BAP trait severity, ranging from subclinical to clinical levels. As such, it is capable of assessing a wide range of trait/symptom severity in both individuals with ASD and typically developing individuals (Ingersoll et al., 2011).

Psychometric limitations. While the BAPQ is frequently used in research and clinical practice, several issues may limit its usefulness. Indeed, some researchers have called into question whether the BAPQ accurately and comprehensively assesses BAP characteristics (Nayar et al., 2018). For example, Nayar and colleagues (2018) suggest that the BAPQ may under classify the BAP compared to direct assessment measures. In their study, they recruited PCAs and compared the proportion of the sample identified as presenting with the BAP (i.e., BAP+) across direct assessments (ADOS, ADI-R, and the Pragmatic Rating Scale-School Age) and the BAPQ. They found that the BAPQ identified fewer PCAs as BAP+ compared to the direct assessment measures. Nayar et al. (2018) also found that correlations between parents' clinical-behavioural symptom scores and their children's scores were more robust when measured by direct assessments. The results of this research suggest that individuals who are BAP+ may not be accurately identified as such if assessed with the BAPQ. Existing direct assessments have undergone extensive development and are considered the gold-standard of ASD symptomatology assessment. In contrast, the BAPQ is relatively newer than many direct assessment measures, and, consequently, research on the BAPQ is still emerging. Considering that the BAPQ is one of the leading BAP measures and is widely used in research and clinical practice (Beaver, 2018), these findings support the need for further research focussed on measure development and psychometric improvements.

One such area requiring additional focus is the internal structure and validity of the BAPQ. The creators of the BAPQ report that it has a three-factor structure with correlated factors (i.e., Aloof, Rigid, and Pragmatic Language; Hurley et al., 2007). As previously mentioned, this proposed factor structure has generally been supported by subsequent research (e.g., Bang et al., 2022; Ingersoll et al., 2011; Sasson et al., 2013a; Wainer et al., 2011); however, the Pragmatic

Language factor has been identified as problematic in some studies. Specifically, a few studies have reported that items on the Pragmatic Language factor fail to demonstrate simple structure due to inadequate item loadings or cross-loadings (Godoy-Gimenez et al., 2018; Kim & Kim, 2022; Lorenz & Algner, 2021; Sasson et al., 2013a). For example, Godoy-Gimenez and colleagues (2018) conducted a confirmatory factor analysis (CFA) to test the three-factor structure (i.e., Rigid, Aloof, and Pragmatic Language), with primary loadings, of a Spanish version of the BAPQ using a sample of 349 undergraduate students (26% male; 73.9% female) and found that factor loadings on the Pragmatic Language factor were low to moderate, with four out of 12 items failing to have adequate loadings (i.e., the items had non-significant loadings for Pragmatic Language). This suggests that these items may not accurately capture pragmatic language abilities. Furthermore, when examining local fit related to factor loadings, half the items on the Pragmatic Language subscale had modification indices greater than 10, indicating local misfit. In regards to global fit, modification indices suggested that the main contributors to global misfit were items on Pragmatic Language. This is in line with earlier research which suggested that the comparative fit index (CFI) of the BAPQ are unacceptable (Broderick et al., 2015). Due to the poor fit associated with the Pragmatic Language factor, Godoy-Gimenez and colleagues (2018) examined an alternative model with Pragmatic Language removed. They found that model fit improved, supporting the use of the Aloof and Rigid factors, but not the Pragmatic Language factor. Considering these findings, Godoy-Gimenez et al. (2018) suggest that it is unadvisable to use Pragmatic Language scores to represent or infer pragmatic language abilities.

Other research has also revealed problematic factor loadings on Pragmatic Language. Indeed, Sasson and colleagues (2013a) attempted to replicate and extend the original research on

the BAPQ conducted by Hurley et al. (2007). Specifically, exploratory factor analysis (EFA) was used to examine the content and structure of the BAPQ subscales in a sample of parents of children with a DSM-IV diagnosis of Autistic Disorder, Asperger's Disorder, or Pervasive Developmental Disorder Not Otherwise Specified, using maximum likelihood discrepancy function and target rotation on both self and informant versions. A three-factor structure was supported, accounting for 45% of the variance for informant ratings and 39% of the variance for self-ratings. However, on the self-report version, one third of the items (items 7, 10, 11, and 21) proposed to measure pragmatic language failed to load onto their respective factor or cross-loaded. Furthermore, its internal consistency, while acceptable, was the lowest of all three subscales (Aloof: $\alpha = 0.92$; Rigid: $\alpha = 0.86$; Pragmatic Language: $\alpha = 0.80$). Taken together, the results of this research suggest that items on the Pragmatic Language subscale may not be measuring perceived pragmatic language abilities as intended. For example, on the self-report version, item 7 ("I am in tune with the other person during conversation"), item 11 ("I feel disconnected or out of sync in conversations with others") have significant loadings on both Pragmatic Language and Aloof, despite being proposed to assess pragmatic language difficulties (Sasson et al., 2013a). Moreover, item 10 ("My voice has a monotone sound to it") has a stronger loading on Aloof than Pragmatic Language despite its inclusion on the Pragmatic Language subscale. This suggests that these items may also be assessing aspects of aloofness rather than primarily pragmatic language. Additionally, item 21 ("I can tell when someone is not interested in what I am saying") fails to load on any factor. Taken together, these findings indicate that these items may not accurately assess pragmatic language (Godoy-Gimenez et al., 2018; Sasson et al., 2013a).

Indeed, in contrast to the face validity purported by BAPQ creators (Hurley et al., 2007),

it appears that several Pragmatic Language items lack face validity. For example, while the creators of the BAPQ suggest that the Pragmatic Language subscale assesses "deficits in the social aspects of language, resulting in difficulties communicating effectively or in holding a fluid, reciprocal conversation" (Hurley et al., 2007, p. 1681), some items appear to reflect executive functioning rather than the social use of language. Executive functions are cognitive processes that support goal-oriented behaviours such as planning, regulating, sequencing, learning, and performing tasks and behaviours (Ahmed & Miller, 2011; Baggetta & Alexander, 2016). These processes consist of sub-processes such as working memory, problem solving, shifting/cognitive flexibility, and inhibitory control (Baggetta & Alexander, 2016; Foy & Mann, 2013), and influence outcomes in a variety of areas such as social, academic, and occupational functioning (Baggetta & Alexander, 2016). While pragmatic language abilities likely involve executive functions (Akbar et al., 2013; Blain-Brière et al., 2014; Hutchison et al., 2020), one could argue that several Pragmatic Language items on the BAPQ better tap aspects of executive functioning than pragmatic language abilities. For example, item 11 which references feelings of disconnection or being "out of sync", is ambiguous and may better reflect self-regulation in conversation than the social use of language. Furthermore, item 32 ("I lose track of my original point when talking to people") may better tap working memory abilities than the social use of language (i.e., pragmatic language). Additionally, item 4 ("It's hard for me to avoid getting sidetracked in conversations") may better reflect aspects of working memory and inhibitory control than pragmatic language. Given that it is well-supported that some individuals with ASD experience impairments in executive functioning (e.g., Hill, 2004), it is plausible that these deficits would extend to subclinical presentations. Thus, it is possible that executive functioning deficits may be a fourth component of the BAP. Indeed, several studies have suggested that

specific cognitive deficits are part of the BAP (e.g., Fombonne et al., 1997; Gökçen et al., 2016). Much of this research has focussed on weak central coherence, social cognition deficits, and executive functioning difficulties, all of which are implicated in the etiology of ASD (Best et al., 2008). The research regarding executive functioning impairments in people with the BAP is mixed and most of the extant research has been concerning relatives of individuals with ASD. For example, some research suggests that first-degree relatives of people with ASD experience executive functioning deficits (Hughes et al., 1997; Piven & Palmer, 1997; Warren et al., 2012), while other studies have failed to replicate these findings (e.g., Losh et al., 200; Pilowsky et al., 2007). The research evidence regarding executive functioning in members of the general population with the BAP is also mixed (e.g., Best et al., 2008; Kunihira et al., 2006). As such, the relationship between executive functioning skills and the BAP is inconclusive and further research is needed to inform whether deficits in executive functioning are a part of the BAP expression.

The aforementioned structural and validity-related issues are undesirable as they make interpretation of scores difficult. The questionable internal structure and construct validity of the BAPQ is concerning considering the wide use of this measure and has implications for its use in both research and clinical settings. Indeed, the BAPQ is increasingly used in a variety of countries and has been translated into several different languages (e.g., Spanish, Chinese, Korean, German, Swedish; Bang et al., 2021; Godoy-Gimenez et al., 2018, Kim & Kim, 2022; Lin et al., 2021; Lorenz & Algner, 2021). Research evaluating the psychometric properties of the BAPQ has primarily been conducted in the United States (e.g., Hurley et al., 2007; Ingersoll et al., 2011; Sasson et al., 2013a) or with versions of the BAPQ that have been translated into other languages. It is uncertain whether the psychometric properties of the BAPQ would be similar

when used with English-speaking Canadians given the impact that culture can have on trait development (e.g., Canadian versus American, individualistic cultures versus collectivist cultures) and perhaps the expression of BAP characteristics.

At present, there is a paucity of research examining the internal structure and construct validity of the BAPQ, and considering that extant findings are mixed, researchers have called for additional examinations of the psychometric properties of the BAPQ. Indeed, the literature base available to inform the English-speaking Canadian context is even more limited. Given the aforementioned problems with item loadings on the Pragmatic Language subscale of the BAPQ, suggesting poor construct validity, and the lack of research focused on examining and improving the psychometric properties of the BAPQ, the present study attempts to address this by examining some sources of evidence for construct validity. The results of the present study may have implications for the use of the BAPQ in research and clinical contexts in both Canada and countries beyond.

Study Objectives

Using a sample of undergraduate students, this study examined the factor structure of the self-report version of the BAPQ. Specifically, this study examined whether the actual internal structure of the self-report version of the BAPQ matches the proposed three-factor structure with primary loadings (Hurley et al., 2007). Second, internal consistency was examined for the total scale as well as each subscale. Third, given that some research has suggested that the Pragmatic Language factor lacks simple structure, this study evaluated the construct validity of the BAPQ, with a particular focus on the Pragmatic Language subscale, to inform whether Pragmatic Language items are accurately measuring pragmatic language or whether they are examining related constructs, such as executive functioning. To examine its construct validity, associations

between Pragmatic Language and several constructs were examined, including self-reports of pragmatic language and executive functioning. Fourth, given that previous research has supported associations between the BAP and the five-factor model of personality, to further examine the construct validity of the BAPQ, associations between the BAPQ and a measure of personality were also examined. Fifth, this study evaluated whether the BAPQ is robust across biological sex to determine whether this measure is assessing the BAP similarly across males and females. To my knowledge, this is the first study examining the psychometric properties of the BAPQ in a Canadian sample and thus serves to inform the English-speaking Canadian context.

Method

Participants

In determining sample size, a minimum of 10 participants per item is required to perform a CFA (MacCallum et al., 1996). Given that the BAPQ has 36 items, a minimum of 360 participants would need to be recruited. However, after taking into account all pairwise comparisons and 36 variances, 666 pieces of information needed to be entered into the model and several parameters needed to be estimated, such as factor loadings, residual variances, and factor correlations. Given this, a sample size of at 720 would be needed to increase the likelihood of successful data fit. Nine hundred undergraduate students attending the University of Manitoba were recruited to take part in this study. Participants were introductory psychology students who signed up to participate in this study for course credit. While university students are a convenience sample, it is also a population that is well-suited to this study's aims as it provides access to a large non-clinical sample. Eight hundred and ninety-nine participants of the 900 recruited were eligible to participate.

Exclusion criteria

Participants were not eligible to participate in this study if they had a prior diagnosis of ASD or a related developmental disorder (e.g., Asperger disorder), a history of severe psychopathology (e.g., schizophrenia), or significant brain injury requiring hospitalization as these conditions may produce effects that could confound the results. Additionally, participants who were not fluent in English (determined by participants' response to a screening question in the online research study portal) were not eligible to participate as this study used an English version of the BAPQ and difficulties communicating in English may act as a potential confound, particularly as this study involves the examination of self-perceived pragmatic language abilities. Participants were screened for these exclusion criteria with pre-screening questions in the online program used for introductory psychology research participation at the University of Manitoba (i.e., SONA System). Participants who answered in the affirmative to these pre-screening questions were filtered out and not eligible to participate. The collection of participant's health information (e.g., mental health diagnoses) allowed for an additional check (e.g., if they indicated they had a diagnosis of ASD, schizophrenia). One participant was excluded as they indicated that they had a diagnosis of ASD and thus did not meet eligibility criteria, resulting in the final sample size of 899 participants. Demographic characteristics of the sample can be found in Table 1.

Study Design and Procedures

This quantitative study utilized a correlational research design in that data was collected to examine the relationships amongst variables of interest using confirmatory factor analysis, correlation, and multiple regression (see Data Analysis below). More specifically, an online survey, consisting of questionnaires intended to measure the variables of interest, was used for data collection.

Participants completed an online questionnaire created using *Qualtrics*, which consisted of demographic questions, the BAPQ (Hurley et al., 2007); the Communication Checklist – Self-report (CC-SR; Bishop et al., 2009), a self-report of pragmatic language abilities; the Amsterdam Executive Functioning Inventory (AEFI; Van der Elst et al., 2012), a self-report of executive functioning; and The Big Five Personality Inventory (BFI; John et al., 1991), a self-report of personality. The online questionnaire did not reference the titles of these measures to decrease the likelihood that titles could influence participants' responses, but instead included abbreviations of the measure names. Due to the COVID-19 pandemic, the online questionnaire was not administered on the University of Manitoba campus and instead was completed remotely from a location of the participants' choosing. Data was collected between June 2021 and December 2021. This research was approved by the University of Manitoba Fort Garry Campus Research Ethics Board.

Measures

Broad Autism Phenotype Questionnaire (BAPQ)

The BAPQ (Hurley et al., 2007) is a 36-item measure intended to assess the BAP. It is available in a self-report version and an informant version; however, the present study only examined the self-report version. It consists of three subscales, including: Aloof, Rigid, and Pragmatic Language. Each subscale is composed of 12 items (Aloof items: 1, 5, 9, 12, 16, 18, 23, 25, 27, 28, 31, and 36; Rigid items: 3, 6, 8, 13, 15, 19, 22, 24, 26, 30, 33, and 35; Pragmatic Language items: 2, 4, 7, 10, 11, 14, 17, 20, 21, 29, 32, and 34) which can be averaged to provide summary scores for each subscale. Summary scores can range from one to six, with higher scores suggesting a higher level of that particular trait (Davidson et al., 2014). A total score can also be computed by averaging all 36 items. For each item, respondents are asked to rate the

extent to which each item applies to them on a 6-point Likert-type scale, ranging from 1 (*Very rarely*) to 6 (*Very often*). Items do not provide a neutral response option, forcing respondents to rate their responses above or below neutral. In an attempt to prevent a response set bias, several items are reverse scored, including 1, 3, 7, 9, 12, 15, 16, 19, 21, 23, 25, 28, 30, 34, and 36. When completing the questionnaire, respondents are asked to base their responses on interactions they have with most people and not just special relationships (e.g., immediate family members) and the way that they have behaved during most of their adult life rather than specific time periods. Additionally, respondents are instructed to take their best guess if unsure how to answer an item. As mentioned, extant research is unclear regarding its psychometric properties. Hurley et al. (2007) suggest that the BAPQ is an efficient and valid measure for identifying individuals with the BAP in both males and females. The creators of the BAPQ suggest that a mean score of 3.15 or higher indicates the presence of the BAP (i.e., BAP +; Hurley et al., 2007). More recently, Sasson et al. (2013) proposed a clinical cut-off for females of 3.17 and for males of 3.55, suggesting that the higher cut-offs would result in fewer false positives. The BAPQ has frequently been used in research using university student samples (e.g., Godoy-Gimenez et al., 2018; Kim & Kim, 2022; Kurtz et al., 2023).

Amsterdam Executive Function Inventory (AEFI)

The AEFI is a 13-item self-report measure of executive functioning, originally intended for use with adolescents (Van der Elst et al., 2012). Items on this scale are rated on a three-point Likert-type scale in which participants rate the extent to which they believe a response is true of them (1 = *Not true*; 2 = *Partly true*, 3 = *True*). It consists of three factors, including: Attention, Self-Control and Self-Monitoring, and Planning and Initiative. More recently, this measure was adapted for use with young adults (Baars et al., 2015) and research suggests that its scores have

acceptable psychometric properties (Van der Elst et al., 2012). Coefficient alphas for the adapted version ranges between 0.65 and 0.78, based on a sample of first-year university students in the Netherlands (Baars et al., 2015).

The Big Five Personality Inventory (BFI)

The BFI is a self-report measure assessing five personality dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. It consists of 44 items and is a relatively brief measure compared to other multidimensional personality inventories, taking approximately five minutes to complete. Respondents rate the extent to which they agree with each item using a 5-point Likert scale, ranging from 1 (*Disagree Strongly*) to 5 (*Agree Strongly*; John et al., 1991). Scores on the BFI are suggested to have acceptable psychometric properties (Arterberry et al., 2014; John et al., 1991) and the BFI has been widely used in social sciences research, including with university student samples. Coefficient alphas for each of the five personality facets are suggested to be as follows: Extraversion (.86), Agreeableness (.79), Conscientiousness (.82), Neuroticism (.87), and Openness (.83; John et al., 2008). Research evidence is in support of using the BFI across various cultures even when translated into languages other than English (John et al., 2008).

Communication Checklist – Self-Report (CC-SR)

The CC-SR (Bishop et al., 2009) is a 70-item self-report measure of communication difficulties intended for use with older children and adults (ages 10 – 80 years). It consists of three subscales, including: Structural Language, Pragmatic Language, and Social Engagement. The present study only used the Pragmatic Language subscale to assess self-perceived pragmatic language abilities. The Pragmatic Language subscale consists of 22 items and measures the degree of respondents' appropriate pragmatic language use (e.g., "I am told I keep going on

about things no one else is interested in" and "I copy what other people say"). Respondents are asked to rate the frequency with which each item corresponds to themselves using the following response options: 0 (*less than once a week, or never)*, 1 (*at least once a week, but not every day)*, 2 (*once or twice a day)*, or 3 (*several times a day, or always)*. The CC-SR has been standardized and normed. Each subscale has a standard score with a mean of 10 and a standard deviation of 3. Moreover, z-scores, scaled scores, and percentiles are available for each subscale. Higher raw scores and lower scaled scores suggest more communication difficulties. This subscale has acceptable psychometric properties, with an alpha coefficient of 0.88, and requires approximately 5 to 8 minutes to complete (D. Bishop, personal communication, July 13, 2020).

Data Analysis

Descriptive statistics were used to examine the characteristics of the sample. To determine whether the actual internal structure of the self-report version of the BAPQ matches the proposed three-factor structure with primary loadings (Hurley et al., 2007), confirmatory factor analysis was used. To examine the internal consistency of the total scale as well as each subscale, coefficient alpha was used. To evaluate the construct validity, specifically convergent validity, of the BAPQ, Pearson correlations and multiple regression analyses were used. Lastly, to determine whether the BAPQ is robust across biological sex, measurement invariance, across multiple levels, was examined (see Results section for more detailed information about the analyses conducted).

Results

Participant Characteristics

BAPQ data were obtained from 899 undergraduate students (M = 21.30 years old, SD = 11.71 years old) in introductory psychology at the University of Manitoba. Sociodemographic

information is included in Table 1. Information was also collected regarding whether participants had been diagnosed with a mental health condition(s) and whether they had a family history of ASD, which can be found in Tables 2 and 3, respectively. The means and standard deviations for the total scores and subscale scores for all measures are reported in Table 4.

BAPQ Psychometrics

Internal Consistency Reliability

Coefficient alpha is the most commonly used reliability estimate in social sciences and behavioural research. Alternative reliability estimates have been proposed, largely within the omega family, which tend to be computed within a factorial analysis framework. These include omega total, omega RT, and omega h. Omega total is the ratio of the true score variance to the total variance. Omega total has been found to perform similarly to alpha, and can be thought of as a more generalized form of alpha (Xiao & Hau, 2023). In contrast, omega RT tends to be more complex than omega total and often results in higher reliability estimates (McNeish, 2018). Finally, omega h evaluates reliability through computing the strength of the general factor while controlling for the influence of group factors (Xiao & Hau, 2023). While alternative reliability estimates exist, it has been cautioned that no single index is appropriate to use with all data types. Despite its wide use being called into question in prior research (e.g., McNeish, 2018), emerging evidence suggests that under certain conditions, alpha is an acceptable choice. More specifically, when data meets the assumption of normality or is not severely non-normal, when the scale is relatively strong (i.e., generally strong factor loadings), and the scale has five or more points, alpha should provide an accurate reliability estimate for continuous scales (Xiao & Hau, 2023). Given that the BAPQ meets the prior considerations, coefficient alpha was selected as a reliability estimate in the present study. Coefficient alpha also has two statistical assumptions

based on classical test theory including that the test is unidimensional and that the items are tau-equivalent. Coefficient alpha is fairly robust to violations of these assumptions when sample size is large, still providing fairly accurate reliability estimations (Edwards et al., 2021). Moreover, simulation studies suggest that even when tau-equivalence is violated, the negative bias on coefficient alpha is minimal and alpha may actually outperform other reliability estimates (except omega) when considering the distance from the population reliability estimate (Edwards et al., 2021). Simulation studies also indicate that the number of items within a scale impacts reliability, with more items serving to protect against a downward-biased coefficient alpha (Green & Yang, 2009). Thus, despite the BAPQ being a multidimensional test (i.e., a test with higher-order factors; a second-order factor model) with correlated factors (and thus not meeting the assumption of unidimensionality) and items on this measure not being tau-equivalent (evident in their differing factor loadings), given the aforementioned reasons, coefficient alpha was used as a reliability estimate in the present study.

Coefficient alpha reliability estimates were examined for the total BAPQ scale score as well as for each of the three subscales using IBM SPSS Version 23. An alpha of 0.80 and above was used as the criterion indicating a sufficient level of internal reliability (Henson, 2001). For the total sample, alpha for the total scale score was 0.90, which indicates a high level of internal consistency. The alpha for the Aloof subscale was 0.88 and for the Rigid subscale was 0.82, both indicating acceptable internal consistency. In contrast, the alpha for the Pragmatic Language subscale was 0.77 which was below the criterion used for indicating sufficient internal consistency.

Reliability estimates were also examined separately for males and females. In terms of the total scale, alpha was 0.90 for females and 0.87 for males, both of which indicated a high

level of internal consistency. The alpha for the Aloof subscale was 0.89 for females and 0.84 for males, indicating acceptable internal consistency. The alpha for the Rigid subscale was 0.82 for females, which suggested good internal consistency. In contrast, for males, the alpha for the Rigid subscale was 0.77, which was just below the threshold for acceptable internal consistency. The alpha for the Pragmatic Language subscale for both males and females was 0.77, which fell below the criterion used for indicating acceptable internal consistency.

Confirmatory Factor Analysis

A confirmatory factor analysis (CFA) was conducted to examine the three-factor structure of the BAPQ using MPlus 8.7 base program with combination add-on (Muthén & Muthén, 1998-2021). All variables were determined to have less than 5% of data missing. The default in MPlus for missing data management is multiple imputation, which was implemented in the present study. I used maximum likelihood estimation for parameter estimation for the CFA (Muthén & Muthén, 1998-2021). Prior to conducting confirmatory factor analytic analyses, statistical assumptions were evaluated. Maximum likelihood estimation relies on several statistical assumptions, including sufficient sample size, the use of indicators that approximate interval-level scales, and multivariate normality (Brown, 2015). Regarding sample size, a sample size of at 720 was determined to be required (see Method) to increase the likelihood of successful data fit. The final sample size meeting eligibility criteria was 899, and thus, exceeded the minimum required sample size of 720. The measures used in this study were all interval-level scales in that the order and the difference between values on the scale hold meaning. Lastly, the assumption of multivariate normality was met evidenced by a visual inspection of Q-Q plots which revealed no significant sagging above or below the line which would have been indicative of kurtosis. Furthermore, quantiles closely followed the diagonal line suggesting multivariate

normality. It is also important to note that with large sample sizes such as those present in this study, maximum likelihood estimations are robust to small deviations in normality, in part due to the Central Limit Theorem (Brown, 2015).

Overall model fit was examined using fit indices. Given that the chi-square statistic is susceptible to the influence of other factors such as sample size (Brown, 2006; Byrne, 2010), it is generally recommended to examine multiple fit indices when determining model-data fit (e.g., Hu & Bentler, 1999). The following fit indices were examined: Chi-Square Goodness of Fit Test, Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Standardized Root Mean Square Residual (SRMR). Models were determined to have acceptable fit to the data if the following criteria were met: non-significant chi-square statistic, RMSEA between < 0.06 and 0.08, CFI and TLI ≥ 0.95, and SRMR ≤ 0.08 (Hu & Bentler, 1999). Four out of five fit indices suggested that the 3-factor model had unacceptable fit including the Chi-Square Goodness of Fit Test, CFI, TLI, and SRMR. Only the RMSEA fit statistic was within the acceptable range; although, it is possible that the chi-square statistic may have been significant due to the large sample size. A summary of measurement model findings is provided in Table 5.

As proposed by the creators of the BAPQ (Hurley et al., 2007), this revealed that the three factors on the BAPQ were correlated with one another. Moderate positive correlations were found between the Aloof factor and the Pragmatic language factor ($r = .60$, $p = .000$), the Aloof factor and the Rigid factor ($r = .58$, $p = .000$), and the Pragmatic language factor and the Rigid factor ($r = .51$, $p = .000$). See Figure 1 for a path diagram of the original model of the BAPQ.

Measurement invariance. Measurement invariance for the BAPQ original model (Model 1) pertaining to biological sex was evaluated next. The chi-square change statistic is

most commonly used to assess measurement invariance (Byrne, 2010); however, given that chi-square is highly influenced by sample size; with large samples, even small differences between groups can lead to a significant chi-square and cause the researcher to (possibly inaccurately) conclude that the model does not have measurement invariance (Byrne, 2010). As such, in the cases of large sample sizes, such as in the present study, it is also recommended to evaluate the CFI change statistic, which is the absolute difference between the more-constrained (more restrictive) model and the less constrained (less restrictive) model. In order to assume measurement invariance using the CFI change statistic, the difference must be <.01 to uphold invariance (Byrne, 2010). Given that this study had a relatively large sample size, the CFI change statistic was used in addition to the chi-square change statistic to assess for measurement invariance across biological sex (i.e., males and females).

First, model fit for males and females was examined, separately. Separate CFAs were conducted for males ($\chi^2(591) = 1343.49$; $p = .000$; *CFI* = 0.66; *TLI* = 0.63; *RMSEA* = 0.08), and females ($\chi^2(591) = 2918.97$; $p = .000$; *CFI* = 0.74; *TLI* = 0.72; *RMSEA* = 0.08). The model for both males and females failed to meet criteria for good fit across most fit statistics, with only the RMSEA value being minimally acceptable. Next, the chi-square and CFI change statistics were compared between the grouped and ungrouped (the entire sample not separated by biological sex) models. Both the chi-square and CFI change statistics upheld configural invariance (information reported in Table 6) and so analysis proceeded to a more restrictive measure of invariance: metric invariance, which constrains factor weights (loadings) to be equal across males and females. While the chi-square change statistic was significant, suggesting differences between males and females, an examination of the CFI change statistic suggested that metric invariance was upheld as the absolute difference was <.01. Given that the chi-square change

statistic was likely impacted by the large sample size used in the present study, metric invariance was assumed based on the CFI change statistic (information reported in Table 7). Thus, analysis proceeded to a more restrictive measure of structural invariance in which the covariances between the factors are constrained to be equal across groups (Byrne, 2010). The chi-square statistic was significant and thus did not provide evidence for structural invariance. In contrast, the CFI change statistic provided sufficient evidence for structural invariance (information reported in Table 8). In sum, analyses yielded sufficient evidence to suggest model invariance between males and females given that every parameter constraint upheld invariance using the CFI change statistic as an indicator. This suggests that males and females respond similarly to items on the BAPQ.

Improving Model Fit. Factor loadings were examined to determine whether specific items may have been contributing to unacceptable data-model fit due to poor factor loadings. Factor loadings can be found in Table 9. Factor loadings < 0.30 were considered weak and factor loadings ≥ 0.60 were considered strong. A total of two items were eliminated as they did not meet a minimum criterion of having a primary factor loading of ≥ 0.30. Both items deleted were on the Pragmatic Language factor including, item 17 (“I have been told that I talk too much about certain topics”), which had a factor loading of .28, and item 21 (“I can tell when someone is not interested in what I am saying”) which had a factor loading of .10. These items were not included in further analyses.

To determine how model fit may be improved, modification indices were generated and their effect on the model fit when implemented was examined. Examination of modification indices is a commonly used practice when models that are supported by sound theory achieve less than acceptable fit (Bryant et al., 1999). Only modifications which made theoretical sense

were implemented. This revealed that allowing for covariance between residual terms associated with item 1 ("I like being around other people") and item 9 ("I enjoy being in social situations"), $\chi^2(1) = 63.304$, $p = <.00001$, would significantly improve model fit. Modification indices also revealed that allowing for covariance between the residual terms associated with item 28 ("I am warm and friendly in my interactions with others") and 12 ("People find it easy to approach me") on the Aloof factor, $\chi^2(1) = 88.075$, $p = < .00001$, would significantly improve model fit. Additionally, allowing for covariance between items 4 ("It's hard for me to avoid getting sidetracked in conversation") and 32 ("I lose track of my original point when talking to people") on the Pragmatic Language factor, $\chi^2(1) = 185.384$, $p = < .00001$, would significantly improve model fit. On the Rigid factor, modification indices revealed that permitting covariance between items 24 ("I act very set in my ways") and 19 ("I look forward to trying new things"), $\chi^2(1) = 81.882$, $p = < .00001$, and items 22 ("I have a hard time dealing with changes in my routine") and 13 ("I feel a strong need for sameness from day to day"), $\chi^2(1) = 81.055$, $p = < .00001$, would significantly improve model fit. These modifications were implemented resulting in a new model (Model 2). See Figure 2 for a path diagram of Model 2.

The same fit statistics were examined to evaluate the overall data-model fit for Model 2, which can be found in Table 10. The chi-square statistic was still significant, likely due to the influence of sample size and the RMSEA was still within the acceptable range. The CFI, TLI, and SRMR values also improved, indicating better model fit; however, were still below the threshold suggested for good data-model fit. While beyond the scope of the present study which was focused on exploring ways that model fit might be improved, this indicates that future research needs to continue to examine construct validity to further improve model fit. Remaining analyses, focussed on convergent validity, are applied to the original model (Model 1).

Validity Based on Relationships of the BAPQ with Measures of Other Variables

Construct validity was evaluated, with a focus on convergent validity (the degree to which constructs that are proposed to be theoretically related are related). Structural equation modelling was originally intended to be used to examine the construct validity of the BAPQ; however, this was not possible given that MPlus yielded a parameter error. As an alternative approach, Pearson correlation coefficients were estimated between total scale and subscale sum-scores of the measures of interest for the total sample (Table 11). Pearson correlation coefficients between total scale and subscale sumscores of the measures of interest for males (Table 12) and females (Table 13) were also computed. Multiple regression analyses (using the enter method) were then used to examine the construct validity of the Pragmatic Language subscale of the BAPQ in greater detail. Only correlations (based on the total sample) of 0.3 or higher between BAPQ Pragmatic Language and other related variables were entered into multiple regression analyses. The results of correlational analyses and multiple regression analyses are detailed below.

Pearson Correlations between BAP Variables and Related Constructs. First, Pearson correlations were used to examine the relationships between the BAPQ total scale and subscale (Rigid, Aloof, and Pragmatic Language) sumscores with theoretically related constructs: executive functioning (as measured by the AEFI) and personality traits (as measured by the BFI).

BAPQ Total Score. Moderate negative correlations were found between the BAPQ total score and several personality traits measured via the BFI, including Extraversion, Agreeableness, and Conscientiousness. A weak negative correlation was found between the BAPQ total score and Openness. These findings suggest that higher BAPQ scores were associated with lower levels of the aforementioned personality traits. The BAPQ total score was moderately and

positively correlated with Neuroticism, suggesting that higher scores on the BAPQ are associated with higher levels of neuroticism. The BAPQ total score had a weak positive correlation with the total score on the AEFI and the Self-Control subscale, and a moderate positive correlation with the Attention subscale. The BAPQ total score had a weak negative correlation with the Planning subscale of the AEFI (see Table 11).

Rigid Subscale. The Rigid subscale of the BAPQ had weak negative correlations with Extraversion, Agreeableness, Conscientiousness, and Openness; it was moderately and positively correlated with Neuroticism. The Rigid subscale also had weak positive correlations with the AEFI total score, Attention subscale, and Self-control subscale; it was unrelated to the Planning subscale of the AEFI.

Aloof Subscale. The Aloof subscale of the BAPQ had a strong negative correlation with Extraversion. It also had weak negative correlations with Agreeableness, Conscientiousness, and Openness. In contrast, it had a weak positive correlation with Neuroticism. The Aloof subscale also had weak positive correlations with the AEFI total score and with the Attention subscale. In contrast, it had a weak negative correlation with the Planning subscale and was not significantly related to the Self-Control subscale.

Pragmatic Language Subscale. To evaluate whether the Pragmatic Language subscale of the BAPQ has convergent validity with another measure of pragmatic language, the Pragmatic Language subscale of the CC-SR, Pearson's correlations were examined. This revealed that the two measures of pragmatic language are strongly and positively correlated with one another, suggesting that it does indeed have convergent validity with another measure of pragmatic language. The Pragmatic Language subscale of the BAPQ was also found to have weak negative correlations with Extraversion and Agreeableness, and a moderate negative correlation with

Conscientiousness. It was moderately and positively correlated with Neuroticism and uncorrelated with Openness. Given that several items on the Pragmatic Language subscale appear that they may also tap into executive functioning, correlations between the Pragmatic Language subscale and the AEFI total score and subscale scores were also examined. This revealed that the Pragmatic Language subscale was moderately and positively correlated with the AEFI total score as well as with the Attention and Self-Control subscales. It had a weak negative correlation with the Planning subscale.

A Closer Look at the Construct Validity of the Pragmatic Language Subscale. Given that the construct validity of the Pragmatic Language subscale on the BAPQ was of particular interest, it was examined in greater detail using multiple regression (enter method). The first multiple regression model explored the influence of the CC-SR Pragmatic Language measure and the AEFI on the BAPQ Pragmatic Language subscale. Given that the Pragmatic Language subscale of the BAPQ was also found to be associated with Extraversion, Neuroticism, and Conscientiousness, in addition to the AEFI, these variables were also entered as predictors in a second model. Prior to implementing multiple regression, the assumptions for multiple regression were evaluated.

Independence of Errors. First, the assumption of independence of errors was assessed using the Durbin-Watson statistic in which values can vary from 0 to 4. A value of two indicates that the residuals are not correlated while values less than 1 and greater than 3 suggest that the residuals may be correlated (Field, 2013). The Durbin-Watson values for both Model 1 (Durbin-Watson = 2.13) and Model 2 (Durbin-Watson = 2.03) were greater than 1 and less than 3, therefore meeting this assumption.

Normality. In terms of the assumption of normality, with large sample sizes, it is not

recommended to use significance tests of normality as these tests are susceptible to influence of large samples and thus can yield significant results for small and unimportant effects (Field, 2013). According to the Central Limit Theorem, as sample sizes increase, the assumption of normality matters less as the distribution is likely to be normal no matter what the sample data looks like (Field, 2013). Given the large sample size in the present study, in line with the Central Limit Theorem, it is assumed that the data is normally distributed. Furthermore, a visual inspection of Q-Q plots suggested that the assumption of normality had been met as there were no major issues with kurtosis evidenced by no significant sagging above or below the line. Additionally, the quantiles generally fell close to the diagonal line, suggesting normally distributed data (Field, 2013). Additionally, the assumption of normally distributed errors had been met as the means of the residuals were close to 0 (Field, 2013).

Homoscedasticity and Linearity. In terms of the assumption of homogeneity of variances, tests such as Levene's test can be used to assess whether the assumption is met. However, like tests of normality, Levene's test is impacted by large sample sizes and thus may yield a significant result when the variances are actually similar. As such, plots of the standardized predicted values against the standardized residuals were examined as an alternative. As the graphs appeared to be a random array of dots, without any prominent funneling or curvatures, the assumptions of homoscedasticity and linearity were met (Field, 2013). Additionally, collinearity and multicollinearity were not present in the data as all tolerance statistics were greater than 0.1 (Field, 2013).

Convergent Validity of the Pragmatic Language Subscale. As mentioned, the first model explored the influence of the CC-SR Pragmatic Language measure and the AEFI on the BAPQ Pragmatic Language subscale. A significant model was found, $F(2, 853) = 349.70$, $p =$

.000, and explained 44.9 % of the variance (R^2_{adj} = .449). More specifically, both CC-SR Pragmatic Language (β = .472, p = .000) and the AEFI (β = .300, p = .000) were found to be significant predictors of BAPQ Pragmatic Language. For the second model, exploring the influence of the AEFI total score, CC-SR Pragmatic Language, Extraversion, Neuroticism, and Conscientiousness on BAPQ Pragmatic Language, a significant model was found, F(5, 824) = 237.69, p = .000, and explained 58.8% of the variance (R^2_{adj} = .588) in BAPQ Pragmatic Language. Additionally, all five variables were significant predictors of BAPQ Pragmatic Language: Conscientiousness (β = -.20, p = .000), Neuroticism (β = .06, p = .024), Extraversion (β = -.27, p = .000), CC-SR Pragmatic Language (β = .378, p = .000), and AEFI (β = .295, p = .000).

Summary of Results

Internal consistency for the original model (Model 1) was examined using coefficient alpha, with an alpha value of 0.80 and above as indicating adequate internal consistency (Henson, 2001). Alpha values exceeded this criterion, indicating good internal consistency, for the BAPQ total score (0.90), the Aloof subscale (0.88) and the Rigid subscale (0.82). The alpha value for the Pragmatic Language subscale was 0.77, suggesting less than sufficient internal consistency.

Confirmatory factor analysis was used to evaluate data-model fit of the proposed three-factor structure with correlated factors of the BAPQ (Hurley et al., 2007). To examine model fit, fit statistics were examined, which suggested poor model fit for four out of five fit indices, including the Chi-Square Goodness of Fit Test, CFI, TLI, and SRMR. Only the RMSEA fit statistic was within the acceptable range. Despite inadequate data-model fit, the results support that the three factors of the BAPQ are indeed correlated with one another.

Measurement invariance for the original model (Model 1) of the BAPQ, across biological sex (i.e., males versus females), was evaluated using the chi-square statistic and the CFI change statistic. Results suggest that measurement invariance, when considering the CFI change statistic, was upheld across all levels, progressing from less restrictive models to more restrictive models. This provides support that males and females respond to the BAPQ in a similar fashion.

Ways in which model fit could be improved was evaluated next. Two items on the Pragmatic Language subscale had inadequate factor loadings and so were deleted. These items included item 17 ("I have been told that I talk too much about certain topics"), which had a factor loading of .28, and item 21 ("I can tell when someone is not interested in what I am saying") which had a factor loading of .10. Modification indices were subsequently examined. This revealed that allowing for covariance between the residuals of several items would improve model fit. The aforementioned item deletions and allowing for covariance between residual terms between particular items, resulted in a second model (Model 2). Fit statistics were examined for Model 2, which revealed improved data-model fit. Indeed, while the chi-square statistic was still significant, the CFI, TLI, and SRMR values improved, albeit still inadequate.

Convergent validity was examined for the original model (Model 1), with a focus on the Pragmatic Language subscale. Pearson correlations and multiple regression analyses were supportive of the convergent validity of the BAPQ as well as the Pragmatic Language subscale in particular. More specifically, the results of multiple regression analyses suggest that the Pragmatic Language subscale of the BAPQ has convergent validity with the CC-SR Pragmatic Language, self-reported executive functioning abilities (as measured by the AEFI), and with several personality traits (i.e., Extraversion, Conscientiousness, and Neuroticism).

Discussion

The BAPQ is the most widely used self-report measure used to assess the BAP in both research and clinical practice. Despite its wide use, a growing body of research provides mixed evidence for its psychometric properties, with some research indicating strong validity and reliability (e.g., Hurley et al., 2007; Ingersoll et al., 2011; Sasson et al., 2013) and other research suggesting that certain psychometric properties of the BAPQ may be inadequate, particularly for the Pragmatic Language subscale (Godoy-Gimenez et al., 2018; Nayar et al., 2018; Sasson et al., 2013a). Given that there has been relatively little research conducted regarding the psychometric properties of the BAPQ, and none to date using a Canadian sample, this study was aimed at extending previous research on the psychometric properties of the BAPQ in a large sample of Canadian undergraduate students. This study was also aimed at providing recommendations for the improvement of the psychometric properties of the BAPQ, and thus, the measurement of the BAP.

Factor Structure and Model Fit

A CFA was implemented to evaluate the psychometric properties of the BAPQ. This revealed that the three factors of the BAPQ are moderately and positively correlated with one another. This is consistent with the factor structure that was proposed by the creators of the BAPQ (Hurley et al., 2007) and prior research (e.g., Ingersoll et al., 2011; Lorenz & Algner, 2021; Sasson et al., 2013a; Wainer et al., 2011). While the correlations between the Pragmatic Language and Aloof factors and the Aloof and Rigid factors in this study were slightly higher than those reported by the creators of the BAPQ (Aloof and Rigid: r = .54, Aloof and Pragmatic Language: r = .53), the correlation between the Rigid and Pragmatic Language factors was the same (r = .51). An examination of fit indices revealed inadequate data-model fit. Specifically, all fit indices except the RMSEA failed to meet the threshold required for good data-model fit,

suggesting questionable factorial validity, which is in line with several existing studies (e.g., Broderick et al., 2015; Kim & Kim, 2022; Lin et al., 2021; Lorenz & Algner, 2021; Godoy-Gimenez et al., 2018). As has been found in prior studies, this study also identified that the main contributors to model misfit are Pragmatic Language items. Indeed, consistent with some prior literature (Godoy-Gimenez et al., 2018; Kim & Kim, 2022; Lorenz & Algner, 2021; Sasson et al., 2013a), this study also found psychometric inadequacies with the Pragmatic Language subscale. Specifically, the Pragmatic Language factor was found to lack simple structure as two items (items 17 and 21) failed to adequately load onto this factor. Similar findings were reported by Sasson and colleagues (2013a) in that item 21 failed to load onto the Pragmatic Language factor; however, they also identified that items 7, 10, and 11 also had problematic factor loadings in that they either failed to load onto the Pragmatic Language factor or cross-loaded. Additionally, Godoy-Gimenez et al. (2018) also found that four out of 12 items (items 10, 17, 21, 32) on the Pragmatic Language factor failed to adequately load. Furthermore, consistent with some extant research, this study revealed that the Pragmatic Language factor had the lowest internal consistency reliability of all three subscales. While previous research identified acceptable internal consistency ($\alpha = 0.80$), the coefficient alpha for the Pragmatic Language subscale in this study (α = 0.77) fell short of the minimum threshold. This provides support for the view that certain items on the BAPQ proposed to assess pragmatic language may not be adequately assessing this construct.

In contrast to the findings of the present study and the existing research discussed above, when the factor structure of the BAPQ was assessed in a Swedish-speaking sample of parents of children with and without ASD, acceptable data-model fit was identified ($\chi^2(591) = 518.495$, p = 0.985; SRMR = 0.088; CFI = 1; Bang et al., 2022) for the proposed three-factor structure.

However, this study had a much smaller sample size which may have increased the risk of a type 2 error. Additionally, the sample used in the aforementioned study also included parents of children with ASD, which resulted in BAPQ total scores for the ASD-parent group that were much higher than those reported in previous studies (Hurley et al., 2007; Sasson et al., 2013a, 2013b; Seidman et al., 2011; Shi et al., 2015) which did not include parents of children with ASD. Bang and colleagues (2022) suggest that this may be indicative of a difference in sample compositions between their sample and those of prior studies rather than reflective of the BAPQ itself. A sample that includes parents of children with ASD who obtained relatively higher BAPQ total scores may also in part account for the difference in model fit findings in which the model may fit better for parents of those with ASD versus a sample of non-clinical undergraduate students which was used in this study.

Measurement Invariance Across Biological Sex

This study found that the BAPQ achieved measurement invariance across biological sex (i.e., males and females) at each level of parameter constraint, according to the CFI change statistic, which is less impacted by sample size than the chi-square statistic. Despite this, the majority of model fit indices evaluated in this study suggested inadequate data-model fit for both males and females. This suggests that while the BAPQ may not be accurately assessing the BAP, at the very least, males and females are responding to items in a similar fashion, suggesting that this questionnaire is a valid measure to use with both males and females.

Suggestions for Improvement

Item 17 and 21 on the Pragmatic Language factor failed to load adequately on their respective factor, which is consistent with prior research (e.g., Sasson et al., 2013a). Consequently, these items were dropped from further analyses in the present study. A number of

modification indices suggested that model fit could be improved if covariance between item residual terms was permitted, including between item 1 (“I like being around other people”) and item 9 (“I enjoy being in social situations”), item 28 (“I am warm and friendly in my interactions with others”) and item 12 (“People find it easy to approach me”), item 4 (“It’s hard for me to avoid getting sidetracked in conversation”) and item 32 (“I lose track of my original point when talking to people”), between items 24 (“I act very set in my ways”) and 19 (“I look forward to trying new things”), and items 22 (“I have a hard time dealing with changes in my routine”) and 13 (“I feel a strong need for sameness from day to day”). This may suggest that these items are redundant. Researchers may also want to consider reverse coding these items to evaluate whether respondents are answering items in a consistent way given that inconsistent responding has negative implications for validity and internal consistency (Jozsa & Morgan, 2017). After items 17 and 21 were dropped and the aforementioned modification indices were implemented in a second model of the BAPQ (Model 2), the fit statistics for Model 2 were superior to that of the original model, albeit still inadequate. As such, revision of item content and greater selectivity in item inclusion may need to be considered to further improve model fit.

Convergent Validity

BAPQ and Personality

This study adds to the body of literature regarding the validity evidence of the BAPQ. Specifically, this study examined the convergent validity of the BAPQ. Similar to previous research, this study found that the BAPQ total score is associated with personality traits conceptualized within The Big Five model of personality as assessed by the BFI (Austin, 2005; De Pauw et al., 2011; Schriber et al., 2014; Wainer et al., 2011; Wakabayashi et al., 2006). Negative correlations were found between the BAPQ total score and Extraversion,

Agreeableness, Conscientiousness, and Openness, suggesting that individuals who score higher on the BAPQ tend to be less extraverted, agreeable, conscientious, and open to new experiences. Similar findings have been reported in prior research. Intuitively, this makes sense given that the BAP subsumes traits such as a lack of interest in social connection, a dislike for and resistance to change, and difficulties with the social aspects of language. This study also found that the BAPQ total score was positively associated with Neuroticism, suggesting those who score higher on the BAPQ will tend to also score higher on measures of neuroticism.

The relationships between the Rigid, Aloof, and Pragmatic Language subscales of the BAPQ with The Big Five personality traits were also examined. This study found a negative correlation between Rigid and the following personality traits: Extraversion, Agreeableness, Conscientiousness, and Openness. This suggests that individuals who tend to be more rigid in their behaviour tend to score lower on measures of extraversion, agreeableness, conscientiousness, and openness. This makes intuitive sense and is consistent with research indicating that individuals with ASD tend to be lower in extraversion, agreeableness, conscientiousness, and openness (Suh et al., 2016). Similarly, Godoy-Gimenez and colleagues (2018) found a low/moderate relationship between Rigid and Openness. A positive correlation was found between Rigid and Neuroticism in the present study, suggesting that individuals who score higher in rigidity also tend to score higher on neuroticism. In terms of the relationship between the Aloof subscale on the BAPQ, similarly, negative correlations were found between Aloof and Extraversion, Agreeableness, Conscientiousness, and Openness, suggesting that individuals who are more aloof tend to be lower in extraversion, agreeableness, conscientiousness, and openness. This is not surprising and is consistent with research conducted with individuals with ASD (e.g., Suh et al., 2016). A similar finding was reported by Godoy-

Gimenez et al. (2018) in that individuals who were less aloof were found to be higher in Extraversion and Agreeableness. The Aloof subscale was positively associated with Neuroticism, suggesting that those who are more aloof tend to score higher on measures of neuroticism. The Pragmatic Language subscale was found to be negatively correlated with Extraversion, Agreeableness, and Conscientiousness. This is inconsistent with some previous research in which a statistically significant relationship between the BAPQ Pragmatic Language subscale and Extraversion and Agreeableness was not found (Godoy-Gimenez et al., 2018). Like the study conducted by Godoy-Gimenez and colleagues (2018), a statistically significant relationship was not found between BAPQ Pragmatic Language and Openness in the current study. This suggests that individuals who have more difficulties with pragmatic language tend to be less extraverted, agreeable, and conscientious. This theoretically makes sense given that social communication difficulties may contribute to individuals being less sociable and interested in cooperating and interacting with others. Additionally, several items assessing pragmatic language difficulties on the BAPQ relate to being "out of touch" with others and losing track of one's train of thought/direction in a conversation and appear to tap executive functioning, which may in part explain the inverse relationship with conscientiousness, which taps into self-regulation (e.g., being organized and well-prepared; Costa & McCrae, 1992). In contrast, a positive correlation was found between Pragmatic Language and Neuroticism, suggesting that individuals who have more difficulty with pragmatic language tend to score higher on measures of neuroticism, aligning with previous research (e.g, Lorenz & Algner, 2021).

BAPQ and Executive Functioning

The relationship between the BAPQ total score and executive functioning as measured by the AEFI was also evaluated. Positive correlations were found between the BAPQ total score and

the AEFI total score and the Attention and Self-Control subscales, suggesting that individuals who score higher on the BAPQ likely have more difficulties with executive functioning, and specifically with attention and self-control. A negative association was found between the BAPQ total score and Planning, suggesting that those who obtain higher scores on the BAPQ tend to have difficulties with planning. These findings are in line with research suggesting that individuals with ASD tend to experience deficits in executive functioning (e.g., Hill, 2004; Robinson et al., 2009); thus, it is not surprising that the same would be found for individuals with subclinical ASD traits. Indeed, prior research has revealed that those with elevated levels of ASD traits have difficulties with executive functions (Christ et al., 2010; Gokcen et al., 2016).

A Closer Look at the Pragmatic Language subscale of the BAPQ

If the Pragmatic Language subscale of the BAPQ is purported to measure self-perceived pragmatic language abilities, then it should have convergent validity with other self-report measures of pragmatic language. As such, the association between the Pragmatic Language subscale of the BAPQ and the CC-SR Pragmatic Language subscale was examined. This revealed a positive correlation between the two measures, thus supporting the convergent validity of the BAPQ Pragmatic Language subscale as a measure of pragmatic language.

Given that the BAPQ was also found to be correlated with other theoretically-related constructs including four out of five of The Big Five personality traits and executive functioning, multiple regression analyses (using the enter method) were also conducted to further examine these relationships. Only correlations of 0.3 or higher between BAPQ Pragmatic Language and other related variables were entered into multiple regression analyses. As such, Agreeableness was not included as a predictor variable given that its correlation with BAPQ pragmatic language did not reach this threshold.

The first multiple regression model included the CC-SR Pragmatic Language subscale and the AEFI total score as predictor variables and the BAPQ Pragmatic Language subscale as the dependent variable. Both executive functioning and CC-SR pragmatic language were significant predictors of pragmatic language as measured by the BAPQ. While executive functioning supports pragmatic language abilities (Akbar et al., 2013; Blain-Brière et al., 2014; Hutchison et al., 2020), it is possible that some of the items on the Pragmatic Language subscale of the BAPQ better measure self-perceived executive functioning abilities compared to pragmatic language abilities. For example, item 11 ("I feel disconnected or "out of sync" in conversations with others") appears to address the degree to which one is able attend to the topic and flow of conversation and self-monitor and self-regulate accordingly as opposed to the social use of language. Similarly, item 32 ("I lose track of my original point when talking to people") would seemingly also be measuring attention, self-monitoring, and self-regulation within conversation, all of which are subsumed under the umbrella of executive functioning. Given that self-reported executive functioning abilities was a significant predictor of BAPQ Pragmatic Language, this may suggest that several BAPQ items purported to assess pragmatic language, may actually be more reflective of self-perceived executive functioning. Moreover, these findings also lend support to the idea that executive functioning deficits may be a fourth trait/symptom category of the BAP (e.g., Fombonne et al., 1997; Gökçen et al., 2016), which makes sense given that individuals with ASD often experience difficulties with executive functioning (e.g., Best et al., 2008, Hill, 2004).

A second multiple regression model was run in which CC-SR Pragmatic Language, the AEFI total score, and three personality traits (Extraversion, Conscientiousness, and Neuroticism) were entered as predictors of BAPQ Pragmatic Language. This model was significant and

explained over half of the variance in BAPQ Pragmatic Language. Additionally, all five predictor variables were significant predictors of BAPQ Pragmatic Language. This suggests that BAPQ Pragmatic Language has convergent validity with the aforementioned constructs.

The relationships between BAPQ Pragmatic Language, personality, and executive functioning, suggest that the Pragmatic Language subscale on the BAPQ may not be a clean measure of self-perceived pragmatic language abilities. However, given that pragmatic language skills recruit executive functions and pragmatic language appears to be associated with certain personality traits, it may still be an appropriate measure of pragmatic language; however, should be considered within the context of its relationship to these related constructs. This issue is not limited to the BAPQ Pragmatic Language subscale and instead is reflective of a broader issue in psychological measurement. When attempting to operationalize and measure latent psychological constructs, overlap between related constructs will inevitably contribute to measurement error or "noise" when attempting to measure a singular construct, such as pragmatic language.

Strengths and Limitations

Strengths

To my knowledge, this is the first study to examine the psychometric properties of the BAPQ in a Canadian sample. Most psychometric research to date has been conducted in the United States (e.g., Hurley et al., 2007; Ingersoll et al., 2011; Sasson et al., 2013a) or with versions of the BAPQ that have been translated into other languages (e.g., Spanish, Chinese, Korean, German, Swedish; Bang et al., 2021; Godoy-Gimenez et al., 2018, Kim & Kim, 2022; Lin et al., 2021; Lorenz & Algner, 2021). Thus, it was unclear whether the BAPQ would retain similar psychometric properties when used with English-speaking Canadian respondents given

the impact that culture can have on trait development (e.g., Canadian versus American, individualistic cultures versus collectivist cultures) and perhaps the expression of BAP characteristics. As such, the present study attempted to inform the English-speaking Canadian context. The results of this study may also help to inform research conducted on the BAPQ in other countries. Moreover, this study helps to elucidate the BAP and its relationships to other constructs (e.g., executive functioning, personality traits). Additionally, while some research has suggested problems with the Pragmatic Language factor, this is the first study to investigate the construct validity of this factor in greater detail by examining whether there is prominent overlap between the Pragmatic Language factor and other constructs, which may provide information in regards to whether items on the Pragmatic Language factor may better tap related constructs (e.g., executive functioning). Further, by examining measurement invariance across males and females, the present study informs whether the BAPQ is robust and valid across these two subgroups. Additionally, the large sample size recruited resulted in a well-powered study, allowing for more accurate conclusions to be drawn based on the sample collected.

Limitations

While the proposed study has a number of strengths, several limitations should be considered. First, while the university student sample provides access to a large non-clinical sample, it may limit the generalizability of findings to non-student samples. Additionally, approximately three quarters of the sample identified as female, which may limit the generalizability of findings to males. Despite this, measurement invariance was upheld across males and females suggesting that males and females respond to items on the BAPQ similarly. Additionally, most participants identified as White in the present study. As such, given that the results of this study are based on a mostly White sample, this may limit the generalizability of

the findings to non-White participants. The sample sizes of other racial subgroups were not large enough to examine measurement invariance so it is unclear whether participants belonging to other racial subgroups would respond to items on the BAPQ similarly to White participants. Furthermore, given that data was collected from participants attending post-secondary education, and thus those having a background of higher education, it is also unclear whether the results of this study would extend to adults with lower levels of education.

Another limitation of this study is that all measures used in this study were self-report, which may contribute to common method error. This makes it difficult to conclude whether the variance in scores is attributable to the construct of interest and the actual predispositions of respondents or biased responding due to the common method (i.e., survey; Podsakoff et al., 2003) used for data collection. While the exclusive use of self-reports is generally considered acceptable with typically developing populations, it is well-accepted that people with ASD tend to experience impaired self-awareness which can contribute to biased responding (Meyer et al., 2006; Montgomery et al., 2008). Likewise, emerging research suggests the same may be present for people with the BAP (Möricke et al., 2016; Sasson et al., 2014; Seidman et al., 2011), albeit to a lesser extent. If self-awareness issues impacted the responses of participants in the present study, this may decrease the accuracy of the data generated by these measures. Moreover, given that inclusion in this study was based on participants' responses to certain self-report questions (e.g., whether they have been diagnosed with ASD), there is a risk that participants may not have accurately responded to these questions and would have been erroneously included in the sample, which may have impacted study findings. Additionally, while participants were encouraged to respond to items in terms of how they have been for most of their adult life and in their interactions with most people, it is possible that the COVID-19 pandemic (which was at its

heights during data collection) and the lifestyle changes (e.g., social distancing) and stress resulting from the same may have impacted participants' responding to questions, particularly those tapping social engagement.

Additionally, while both performance-based measures and rating scales of executive functioning are intended to measure the same construct and are commonly used in research and clinical practice, emerging research suggests that they may measure different constructs (Toplak et al., 2013). Performance-based measures of executive functioning are administered under highly standardized conditions so that it can be inferred that (ideally) each examinee has completed the assessment under the same conditions. In contrast, rating measures of executive functioning were developed to provide an ecologically valid perspective on everyday situations that involve executive functions (Roth et al., 2005). If these two measurement formats assess the same construct, then they should be strongly associated; however, research indicates that they are not and instead assess different forms of cognitive and behavioural functioning (Biederman et al., 2008; Toplak et al., 2013). This speaks to the difference between typical and maximal performance, with rating scales measuring typical performance and performance-based measures assessing maximal performance. While rating scales have been suggested to provide a more ecologically valid indication of everyday performance, they are subject to response biases. Considering that the AEFI is a self-report measure, it can only be interpreted as respondents' self-perceived day-to-day executive functioning which may or may not be an accurate indication of their true executive functioning.

An additional limitation of this study pertains to the way in which the CC-SR Pragmatic Language subscale was included in this study's online questionnaire. Specifically, on the CC-SR, Pragmatic Language items are presented interleaved with other items; however, because this

study only used the Pragmatic Language items as a standalone measure of pragmatic language skills, it is unclear whether responses to these items may have been impacted by presenting them in a separate block (D. Bishop, personal communication, July 13, 2020). Moreover, because data was collected using an online questionnaire, participants were able to complete the survey in the location of their choosing, which may have increased the risk of measurement error (e.g., distraction, survey was not completed independently) compared to a more tightly controlled research setting at the University of Manitoba.

Additionally, due to the warning error generated by MPlus when attempting to conduct structural equation modelling to examine the construct validity of the BAPQ with related constructs (e.g., personality, executive functioning), an alternative data analysis approach had to be implemented. This alternative approach involved conducting Pearson's correlations and multiple regressions. While this is an acceptable approach to examine the relationships between theoretically related constructs, conducting several correlations at once may have increased the risk of making a type I error. Despite this, the relationships found between constructs in this study were generally consistent with those reported in prior research.

Lastly, although the present study was aimed in part at elucidating the factor structure of the BAPQ through conducting a CFA, the precise factor structure remains unclear. Despite these limitations, this study provided information about the psychometric properties of the BAPQ and potential changes to the measure to improve its psychometric properties.

Future Directions

Given that this is the first study to examine the factor structure and psychometric properties of the BAPQ in a Canadian English-speaking sample, replication is needed to determine if similar findings are yielded. Furthermore, this study only examined the self-report

version of the BAPQ. Further research evaluating the factor structure and psychometric properties of the informant version of the BAPQ in an English-speaking Canadian sample is needed as well as in other countries and languages. Additionally, considering that the present study used a university-based sample, researchers may consider examining the factor structure and psychometric properties of the BAPQ in a community sample and/or with participants with lower levels of educational attainment. It will also be important for future researchers who are intent on examining the psychometric properties of the BAPQ to recruit participants from the general population using a random sampling approach. This will further help to ensure that the sample is representative of the general population, and, thus, increasing the likelihood that the research results will generalizable to the general population. Moreover, attaining a sample with equal numbers of males and females to further examine measurement invariance of the BAPQ will be important for generating additional evidence that the BAPQ can be used with confidence with both males and females. Future research may also consider examining measurement invariance across other diversity factors (e.g., age, culture).

Furthermore, inadequate model fit and factor loadings for certain Pragmatic Language items suggests that there may be problems with the operationalization of the BAP, and Pragmatic Language in particular, on the BAPQ. The present study attempted to improve model fit through dropping items with inadequate loadings and allowing for covariance between the residual terms of several items. Future researchers examining the factor structure and psychometric properties of the BAPQ may consider addressing the operationalization of the BAP and Pragmatic Language more in depth via revising items with a focus on reducing redundancy, revising Pragmatic Language items with poor loadings and/or adding new items that may better tap into pragmatic language abilities. A subsequent step would then be to implement a CFA for

validation of the revised BAPQ.

As executive functioning was a significant predictor of BAPQ Pragmatic Language and executive functioning difficulties have been theorized to be an aspect of the BAP, in addition to the revision of Pragmatic Language items, future research may explore whether adding a fourth factor assessing executive functioning difficulties improves model fit. Concerning validity evidences for the BAPQ, future researchers may also consider including a performance-based measure of executive functioning in addition to a self-report measure to examine whether similar results are found.

Implications and Conclusions

Over the last few decades, there has been increasing recognition that symptoms of ASD exist on a spectrum, varying in presentation, type, and severity. The dimensional nature of its severity suggests that these symptoms vary from a clinical diagnosis of ASD to subclinical traits present in those without a diagnosis of ASD. This subclinical presentation has become known as the BAP and can present in both relatives of individuals with ASD and the general population. Consequently, there has been increasing interest in studying the BAP, how it impacts various areas of functioning, and how this construct may be utilized in ASD research to assist in elucidating ASD etiology. Various BAP measures have been developed; however, the BAPQ is considered one of the leading BAP questionnaires and is frequently used in research and clinical practice. Despite this, there has been little research examining the psychometric properties of the BAPQ and existing research on its internal structure and construct validity are mixed. Considering that the psychometric quality of psychological measures has implications for scoring and interpretation, more research investigating the psychometric properties of the BAPQ is needed. This will help to ensure that the conclusions being drawn in research and the decisions

being made in clinical practice regarding the BAP are based on accurate assessments. This study provided support for the convergent validity of this measure as the subscales of the BAPQ were appropriately associated with related constructs. However, similar to prior research, the present study did not find support for acceptable model fit, lending support to the idea that it's proposed three-factor structure may be inappropriate. Moreover, this study adds to the growing body of research suggesting that the Pragmatic Language factor in particular may be problematic and that item content and inclusion in the measure may need to be re-evaluated. Overall, the information gained through conducting this study will contribute to a better understanding of the factor structure of the BAPQ in an English-speaking Canadian sample. This study also informed potential modifications to the BAPQ that may improve model fit. Improving the psychometric properties of this measure is paramount in ensuring that researchers and practitioners are accurately measuring the BAP, and that users of the BAPQ feel confident that they are using a measure with empirical support.

References

Adam, D. (2013). Mental health: On the spectrum. *Nature (London), 496*(7446), 416–418. doi: 10.1038/496416a

Ahmed, F. S., & Miller, L. (2011). Executive function mechanisms of theory of mind. *Journal of Autism and Developmental Disorders, 41*, 667-678. doi: 10.1007/s10803-010-1087-7

Akbar, M., Loomis, R., & Paul, R. (2013). The interplay of language on executive functions in children with ASD. *Research in Autism Spectrum Disorders, 7*(3), 494-501. doi: 10.1016/j-rasd.2012.09.001

Almasy, L., Blangero, J., & Leal, S. (2001). Endophenotypes as quantitative risk factors for psychiatric disease: Rationale and Study Design. *American Journal of Medical Genetics, 105*(1), 42-44. doi: 10.1002/1096-8628(20010108)105:1<42::AID-AJMG1-55>3.0.CO;2-9

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Arterberry, B., Martens, M., Cadigan, J., & Rohrer, D. (2014). Application of Generalizability Theory to the Big Five Inventory. *Personality and Individual Differences*, *69*, 98–103. https://doi.org/10.1016/j.paid.2014.05.015

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. doi: 10.1016/j.paid.2004.04.022

Baars, M. A., Nije Bijvank, M., Tonnaer, G. H., & Jolles, J. (2015). Self-report measures of executive functioning are a determinant of academic performance in first-year students at a university of applied sciences. *Frontiers in psychology, 6*, 1131. doi: 10.3389/fpsyg.2015.01131

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain, and Education, 10*(1), 10-33.

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., et al. (1995). Autism as a strongly genetic disorder: Evidence from a British twin study. *Psychological Medicine, 25*(1), 63-77.

Bailey, A., Palferman, S., Heavey, L., & Le Couteur, A. (1998). Autism: The phenotype in relatives. *Journal of Autism and Developmental Disorders, 28*(5), 369-392.

Bang, P., Stromberg, M., Meera, S. S., & Igelstrom, K. (2022). Brief report: The Broad Autism Phenotype in Swedish Parents of Children with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders, 52*(10), 4575-4582. doi: 10.1007/s10803-021-05302-3

Baron-Cohen, S., & Wheelwright, S. (1999). "Obsessions" in children with autism or Asperger syndrome: Content analysis in terms of core domains of cognition. *The British Journal of Psychiatry, 175*(5), 484-490. doi: 10.1192/bjp.175.5.484

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,* 5-17. Retrieved from https://www.ncbi.nlm.gov/pubmed/11439754

Beaver, A. G. (2018). Brain Activity in College Students with the Broad Autism Phenotype (Unpublished doctoral dissertation). Stephen F. Austin State University, Texas.

Bellini, S. (2004). Social skills deficits and anxiety in high-functioning adolescents with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities, 19*(2), 78-86. doi: 10.1177/1088357604019002

Best, C., Moffat, V., Power, M., Owens, D., & Johnstone, E. (2008). The Boundaries of the Cognitive Phenotype of Autism: Theory of Mind, Central Coherence and Ambiguous Figure Perception in Young People with Autistic Traits. *Journal of Autism and Developmental Disorders, 38*(5), 840–847. doi: 10.1007/s10803-007-0451-8

Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. *Brain Research, 1380,* 42-77. doi: 10.1016/j.brainres.2010.11.078

Bialosiewicz, S., Murphy, K., & Berry, T. (2013). *An introduction to measurement invariance testing: Resources packet for participants.* Retrieved from http://comm.eval.org/HigherLogic/System?DownloadDocumentFile.ashx?DocumentFileKey=63758fed-a490-43f2-8862-2de0217a08b8

Biederman, J., Petty, C., Fried, R., Black, S., Faneuil, A., Doyle, A., Seidman, L., & Faraone, S. (2008). Discordance between psychometric testing and questionnaire-based definitions of executive function deficits in individuals with ADHD. *Journal of Attention Disorders, 12*(1), 92–102. doi: 10.1177/1087054707305111

Bishop, D., Mayberry, M., Maley, A., Wong, D., Hill, W., & Hallmayer, J. (2004). Using self-report to identify the broad phenotype in parents of children with autistic spectrum

disorders: a study using the Autism-Spectrum Quotient. *Journal of Child Psychology and Psychiatry, 45*(8), 1431–1436. doi: 10.1111/j.1469-7610.2004.00325.x

Bishop, D., Whitehouse, A., & Sharp, M. (2009). *The communication checklist self-report.* Manual. New York: Pearson Education.

Blain-Brière, B., Bouchard, C., & Bigras, N. (2014). The role of executive functions in the pragmatic skills of children age 4-5. *Frontiers in Psychology, 5*, 1-14. doi: 10.3389/fpsyg.2014.00240

Broderick, N., Wade, J. L., Meyer, J. P., Hull, M., & Reeve, R. E. (2015). Model invariance across genders of the broad autism phenotype questionnaire. *Journal of autism and developmental disorders*, *45*(10), 3133-3147. doi: 10.1007/s10803-015-2472-z

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* The Guilford Press.

Brown, T.A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Byrne, B. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York: Routledge.

Chan, D. (2011). Advances in analytical strategies. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (Vol. 1, pp. 85–113). Washington, DC: American Psychological Association. doi:10.1037/12169-004

Chaste, P., & Leboyer, M. (2012). Autism risk factors: genes, environment, and gene-environment interactions. *Dialogues in Clinical Neuroscience, 14*(3), 281–292. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3513682/

Christ, S. E., Kanne, S. M., & Reiersen, A. M. (2010). Executive Function in Individuals with

Subthreshold Autism Traits. *Neuropsychology*, *24*(5), 590–598. https://doi.org/10.1037/a0019176

Colvert, E., Tick, B., McEwen, F., Stewart, C., Curran, S. R., Woodhouse, E., Gillan, N., Hallett, V., Lietz, S., Garnett, T., Ronald, A., Plomin, R., Rijsdijk, Früh-ling., Happé, F., & Bolton, P. (2015). Heritability of autism spectrum disorder in a UK population-based twin sample. *JAMA Psychiatry, 72*(5), 415-423. doi: 10.1001/jamapsychiatry.2014.3028

Constantino, J., Gruber, C., Davis, S., Hayes, S., Passanante, N., & Przybeck, T. (2004). The factor structure of autistic traits. *Journal of Child Psychology and Psychiatry, 45*(4), 719–726. doi: 10.1111/j.1469-7610.2004.00266.x

Constantino, J., & Todd, R. (2003). Autistic Traits in the General Population: A Twin Study. *Archives of General Psychiatry*, *60*(5), 524–530. doi: 10.1001/archpsyc.60.5.524

Costa, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders, 6*(4), 343–359. doi: 10.1521/pedi.1992.6.4.343

Craddock, N., & Owen, M. J. (2010). The Kraepelinian dichotomy – going, going…but still not gone. *The British Journal of Psychiatry, 196,* 92-95. doi: 10.1192/bjp/bp.109.073429

Davidson, J., Goin-Kochel, R. P., Green-Snyder, L. A., Hundley, R. J., Warren, Z., & Peters, S. U. (2014). Expression of the broad autism phenotype in simplex autism families from the Simons Simplex Collection. *Journal of Autism and Developmental Disorders, 44*(10), 2392-2399. doi: 10.1007/s10803-012-1492-1

Dawson, G., Estes, A., Munson, J., Schellenberg, G., Bernier, R., & Abbott, R. (2007). Quantitative assessment of autism symptom-related traits in probands and parents: Broader phenotype autism symptom scale. *Journal of Autism and Developmental Disorders, 37*(3), 523-536. doi: 10.1007/s10803-006-0182-2

Deisinger, J. A., & Rotatori, A. (2015). *The broad autism phenotype*. Bingley, England: Emerald.

De Pauw, S., Mervielde, I., Van Leeuwen, K., & De Clercq, B. (2011). How Temperament and Personality Contribute to the Maladjustment of Children With Autism. *Journal of Autism and Developmental Disorders*, *41*(2), 196–212. doi: 10.1007/s10803-010-1043-6

Edwards, A. A., Joyner, K. J., & Schatschneider, C. (2021). A Simulation Study on the Performance of Different Reliability Estimation Methods. *Educational and Psychological Measurement*, *81*(6), 1089–1117. https://doi.org/10.1177/0013164421994184

Field, A. (2013). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock "n" roll* (4th ed.). Sage.

Folstein, S., & Rutter, M. (1977). Infantile autism: a genetic study of 21 twin pairs. *Journal of Child Psychology and Psychiatry, 18*(4), 297-321.

Fombonne, E., Bolton, P., Prior, J., Jordan, H., & Rutter, M. (1997). A Family Study of Autism: Cognitive Patterns and Levels in Parents and Siblings. *Journal of Child Psychology and Psychiatry*, *38*(6), 667–683. doi: 10.1111/j.1469-7610.1997.tb01694.x

Foy, J. G., & Mann, V. A. (2013). Executive function and early reading skills. *Reading and Writing: An Interdisciplinary Journal, 26,* 453-472.

Furr, R. M., & Bacharach, V. R. (2014). Psychometrics: An introduction (2nd ed.). Sage Publications.

Gerdts, J., & Bernier, R. (2011). The broader autism phenotype and its implications on the etiology and treatment of autism spectrum disorders. *Autism Research and Treatment, 2011,* 545901. doi: 10.1155/2011/545901

Gershon, E., & Goldin, L. (1986). Clinical methods in psychiatric genetics: I. Robustness of genetic marker investigative strategies. *Acta Psychiatrica Scandinavica, 74*(2), 113-118. doi: 10.1111/j.1600-0447.1986.tb10594.x

Geschwind, D. H. (2011). Genetics of autism spectrum disorders. *Trends in Cognitive Sciences, 15*, 409-416. doi: 10.1016/j.tics.2011.07.003

Godoy-Giménez, M., González-Rodríguez, A., Cañadas, F., Estévez, A., & Sayans-Jiménez, P. (2018). Psychometric Properties of the Spanish Version of the Broad Autism Phenotype Questionnaire: Strengths, Weaknesses, and Future Improvements. *Journal of Autism and Developmental Disorders*, *48*(3), 770–783. doi: 10.1007/s10803-017-3438-0

Gökçen, E., Frederickson, N., & Petrides, K. (2016). Theory of Mind and Executive Control Deficits in Typically Developing Adults and Adolescents with High Levels of Autism Traits. *Journal of Autism and Developmental Disorders*, *46*(6), 2072–2087. doi: 10.1007/s10803-016-2735-3

Goldberg, W., Jarvis, K., Osann, K., Laulhere, T., Straub, C., Thomas, E.,…Spence, M. (2005). Brief report: Early social communication behaviors in the younger siblings of children with autism. *Journal of Autism and Developmental Disorders, 35*(5), 657-664. doi: 10.1007/s10803-005-0009-6

Gottesman, I., & Gould, T. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry, 160*(4), 636-645. doi: 10.1176/appi.ajp.160.4.636

Green, S. B., & Yang, Y. (2009). Commentary on Coefficient Alpha: A Cautionary Tale. *Psychometrika, 74*(1), 121–135. https://doi.org/10.1007/s11336-008-9098-4

Gurbuz, E., Hanley, M., & Riby, D. M. (2019). University students with autism: The social and academic experiences of university in the UK. *Journal of Autism and Developmental Disorders, 49*(2), 617-631. doi: 10.1007/s10803-018-3741-4

Gurney, J., McPheeters, M., & Davis, M. (2006). Parental Report of Health Conditions and Health Care Use Among Children With and Without Autism: National Survey of Children's Health. *Archives of Pediatrics & Adolescent Medicine, 160*(8), 825–830. doi: 10.1001/archpedi.160.8.825

Haas, K., Costley, D., Falkmer, M., Richdale, A., Sofronoff, K., & Falkmer, T. (2016). Factors influencing the research participation of adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 46*(5), 1793-1805. doi: 10.1007/s10803-016-2708-6

Happé, F., & Frith, U. (2020). Annual Research Review: Looking back to look forward – changes in the concept of autism and implications for future research. *The Journal of Child Psychology and Psychiatry, 61*(3), 218-232. doi: 10.1111/jcpp.13176

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*(3), 177-189. Doi: 10.10180/07481756.2002.12069034

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental*

Review, *24*(2), 189–233. https://doi.org/10.1016/j.dr.2004.01.001

Hoekstra, R., Bartels, M., Cath, D., & Boomsma, D. (2008). Factor Structure, Reliability and Criterion Validity of the Autism-Spectrum Quotient (AQ): A Study in Dutch Population and Patient Groups. *Journal of Autism and Developmental Disorders*, *38*(8), 1555–1566. doi: 10.1007/s10803-008-0538-x

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55. doi: 10.1080/10705519909540118

Hudepohl, M., Robins, D., King, T., & Henrich, C. (2013). The role of emotion perception in adaptive functioning of people with autism spectrum disorders. *Autism: The International Journal of Research and Practice*, *19*(1), 107–112. doi: 10.1177/1362361313512725

Hughes, C., Leboyer, R, M., & Bouvard, M. (1997). Executive function in parents of children with autism. *Psychological Medicine*, *27*(1), 209–220. doi: 10.1017/S0033291796004308

Hurley, R., Losh, M., Parlier, M., Reznick, J., & Piven, J. (2007). The Broad Autism Phenotype Questionnaire. *Journal of Autism and Developmental Disorders*, *37*(9), 1679–1690. doi: 10.1007/s10803-006-0299-3

Hutchison, S. M., Müller, U., & Iarocci, G. (2020). Parent reports of executive function associated with functional communication and conversational skills among school age children with and without autism spectrum disorder. *Journal of autism and developmental disorders*, *50*(6), 2019-2029. doi: 10.1007/s10803-019-03958-6

Iacono, W. G. (2018). Endophenotypes in psychiatric disease: Prospects and challenges. *Genome*

Medicine, 10(1), 11-11. doi: 10.1186/s13073-018-0526-5

Ingersoll, B., Hopwood, C., Wainer, A., & Brent Donnellan, M. (2011). A comparison of three self-report measures of the broader autism phenotype in a non-clinical sample. *Journal of Autism and Developmental Disorders*, *41*(12), 1646–1657. doi: 10.1007/s10803-011-1192-2

Ingersoll, B., & Wainer, A. (2014). The broader autism phenotype. In F. R. Volkmar, S. J. Rodgers, R. Paul., & K. A. Pelphrey (Eds.), *Handbook of Autism and Pervasive Developmental Disorders*, (4th ed., pp. 28-56)

Jeste, S. S., & Geschwind, D. H. (2014). Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nature Reviews Neurology, 10,* 74-81. doi: 10.1038/nrneurol.2013.278

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The big five inventory- Versions 4a and 54.* Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, *3*(2), 114-158.

Jones, M., & Szatmari, P. (2002). A risk-factor model of epistatic interaction, focusing on autism. *American Journal of Medical Genetics, 114*(5), 558-565. doi: 10.1002/ajmg.10513

Józsa, K., & Morgan, G. A. (2017). Reversed items in Likert scales: Filtering out invalid responders. *Journal of Psychological and Educational Research (JPER), XXV*(1), 7–25.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous child*, *2*(3), 217-250. Retrieved from http://mail.neurodiversity.com/library_kanner_1943.pdf

Kim, H., & Kim, K. (2022). Psychometric properties of the Korean version of the Broad Autism Phenotype questionnaire. *Social Behaviour and Personality, 50*(2), e11153. doi: 10.2224/sbp.11153

Kunihira, Y., Senju, A., Dairoku, H., Wakabayashi, A., & Hasegawa, T. (2006). "Autistic" Traits in Non-Autistic Japanese Populations: Relationships with Personality Traits and Cognitive Ability. *Journal of Autism and Developmental Disorders, 36*(4), 553–566. doi: 10.1007/s10803-006-0094-1

Kurtz, M. R., Kana, R. K., Rivera, D. L., & Newman, S. D. (2023). The role of the broader autism phenotype in anxiety and depression in college-aged adults. *Frontiers in Psychiatry, 14*, 1187298.

Lainhart, J. E., Ozonoff, S., Coon, H., Krasny, L., Dinh, E., Nice, J., & Mcmahon, W. (2003). Autism, regression, and the broader autism phenotype. *American Journal of Medical Genetics, 113*(3), 231-237. doi: 10.1002/ajmg.10615

Landry, O., & Chouinard, P. A. (2016). Why we should study the broader autism phenotype in typically developing populations. *Journal of Cognition and Development, 17*, 584-595. doi: 10.1080/15248372.2016.1200046

Lin, J., Fu, X., Dai, S., Zeng, Y., Yang, R., Liu, R., Yang, J., Shen, Y., & Ou, J. (2021). The broad autism phenotype questionnaire (BAPQ): Strengths, weaknesses, and future improvements in Chinese version. *Asian Journal of Psychiatry, 60*, 102655-102655. doi: 10.1016/j.ajp.2021.102655

Lorenz, T., & Algner, M. (2021). Lost in translation or just too pragmatic?: Validation of the German Broader Autism Phenotype Questionnaire (BAPQ-G0. *Psychological Test Adaptation and Development, 2*(1), 124-135. doi: 10.1027/2698-1866/a000016

Losh, M., Adolphs, R., & Piven, J. (2011). The broad autism phenotype. In D. G. Amaral, G. Dawson, & D. H. Geschwind (Eds.), *Autism spectrum disorders* (pp. 457 – 476). New York, NY: Oxford University Press, Inc.

Losh, M., Childress, D., Lam, K., & Piven, J. (2008). Defining key features of the broad autism phenotype: A comparison across parents of multiple- and single-incidence autism families. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *147*(4), 424-433. doi: 10.1002/ajmg.b.30612

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for convenience structure modeling. *Psychological Methods*, 1(2), 130-149. doi: 10.1037/1082-989X.1.2.130

Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A., & Skuse, D. (2012). Sex differences in autism spectrum disorder: Evidence from a large sample of children and adolescents. *Journal of Autism and Developmental Disorders, 42*(7), 1304-1313. doi: 10.1007/s10803-011-1356-0

Marshall, C., Noor, A., Vincent, J., Lionel, A., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., Thiruvahindrapduram, B., Fiebig, A., Schreiber, S., Friedman, J., Ketelaars, C., Vos, Y., Ficicioglu, C., Kirkpatrick, S., Nicolson, R., … Scherer, S. (2008). Structural Variation of Chromosomes in Autism Spectrum Disorder. *American Journal of Human Genetics*, *82*(2), 477–488. doi: 10.1016/j.ajhg.2007.12.009

Massrali, A., Brunel, H., Hannon, E., Wong, C., iPSYCH-MINERvA Epigenetics Group, Baron-Cohen, S., & Warrier, V. (2018). Integrated genetic and methylomic analyses identify shared biology between autism and autistic traits. *Molecular Autism, 10*(1),

31. doi: 10.1186/s13229-019-0279-z

McNeish, D. (2018). Thanks Coefficient Alpha, We'll Take It From Here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Meyer, J. A., Mundy, P. C., Van Hecke, A. V., & Durocher, J. S. (2006). Social attribution processes and comorbid psychiatric symptoms in children with Asperger syndrome. *Autism: The International Journal of Research and Practice*, *10*(4), 383–402. doi: 10.1177/1362361306064435

Micali, N., Chakrabarti, S., & Fombonne, E. (2004). The broad autism phenotype: Findings from an epidemiological survey. *Autism, 8*(1), 21-37. doi: 10.1177/1362361304040636

Milfont, T. L., & Fischer, R. (2015). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research,* 3, 111–130. doi:10.21500/20112084.857

Montgomery, J. M., Schwean, V. L., Burt, J. A. G., Dyke, D. I., Thorne, K. J., Hindes, Y. L., …& Kohut, C. S. (2008) Emotional intelligence and resiliency in young adults with Asperger's disorder: Challenges and opportunities. *Canadian Journal of School Psychology*, *23*(1), 70-93. doi: 10.1177/0829573508316594

Möricke, E., Buitelaar, J., & Rommelse, N. (2015). Do We Need Multiple Informants When Assessing Autistic Traits? The Degree of Report Bias on Offspring, Self, and Spouse Ratings. *Journal of Autism and Developmental Disorders*, *46*(1), 164–175. doi: 10.1007/s10803-015-2562-y

Murphy, M., Bolton, P. F., Pickles, A., Fombonne, E., Piven, J., & Rutter, M. (2000). Personality traits of the relatives of autistic probands. *Psychological Medicine, 30*(6), 1411-1424.

Muthén, L.L., & Muthén, B.O. (1998-2021). MPlus User's Guide. Eigth Edition. Los Angeles, CA: Muthén & Muthén.

National Institute of Mental Health. (2013). Transforming Diagnosis. Retrieved from https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/transforming-diagnosis.shtml#:~:text=Patients%20with%20mental%20disorders%20deserve,for%20a%20new%20classification%20system.

Nayar, K., Maltman, N., Patel, S. P., Bush, L., Crawford, S., Martin, G. E., & Losh, M. (2018, May). *Measuring the Broad Autism Phenotype in Parents of Children with Autism.* International Society for Autism Research (INSAR). Rotterdam, Netherlands.

Pickles, A., Starr, E., Kazak, S., Bolton, P., Panikolalou, K., Bailey, A.,…& Rutter, M. (2000). Variable expression of the autism broader phenotype: Findings from extended pedigrees. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(4), 491-502. doi: 10.1111/1469-7610.00634

Pilowsky, T., Yirmiya, N., Shalev, R., & Gross-Tsur, V. (2003). Language abilities of siblings of children with autism. *Journal of Child Psychology and Psychiatry, 44*(6), 914-925. doi: 10.111/1469-7610.00175

Pinto, D., Pagnamenta, A., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T., Correia, C., Abrahams, B., Almeida, J., Bacchelli, E., Bader, G., Bailey, A., Baird, G., Battaglia, A., Berney, T., Bolshakova, N., Bölte, S., … Brian, J. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature, 466*(7304), 368–372. doi: 10.1038/nature09146

Piven, J. (2001). The broad autism phenotype: A complementary strategy for molecular genetic studies of autism. *American Journal of Medical Genetics (Neuropsychiatric*

Genetics), 105, 34-35. doi: 10.1002/1096-8628(20010108)

Piven, J., Gayle, J., Chase, G. A., Fink, B., Landa, R., Wzorek, M. M., & Folstein, S. E. (1990). A family history study of neuropsychiatric disorders in the adult siblings of autistic individuals. *Journal of the American Academy of Child and Adolescent Psychiatry, 29*(2), 177-183. doi: 10.1097/00004583-199003000-00004

Piven, J., & Palmer, P. (1997). Cognitive Deficits in Parents from Multiple-incidence Autism Families. *Journal of Child Psychology and Psychiatry*, *38*(8), 1011–1021. doi: 10.1111/j.1469-7610.1997.tb01618.x

Piven, J., Palmer, P., Landa, R., Santangelo, S., Jacobi, D., & Childress, D. (1997). Personality and language characteristics in parents from multiple-incidence autism families. *American Journal of Medical Genetics, 74*(4), 398-411.

Pizzo, L., Jensen, M., Polyak, A., Rosenfeld, J., Mannik, K., Krishnan, A., McCready, E., Pichon, O., Le Caignec, C., Van Dijck, A., Pope, K., Voorhoeve, E., Yoon, J., Stankiewicz, P., Cheung, S., Pazuchanics, D., Huber, E., Kumar, V., Kember, R., … Galesi, O. (2018). Rare variants in the genetic background modulate cognitive and developmental phenotypes in individuals carrying disease-associated variants. *Genetics in Medicine, 21*(4), 816–825. doi: 10.1038/s41436-018-0266-3

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioural research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903. doi: 10.1037/0021-9010.88.5.879

Risch, N., Spiker, D., Lotspeich, L., Nouri, N., Hallmayer, J., …Myers, R. (1999). A genomic screen of autism: Evidence for a multilocus etiology. *The American Journal of Human*

Genetics, 65(2), 493-507. doi: 10.1086/302497

Robinson, E. B., Koenen, K. C., McCormick, M. C., Munir, K., Hallett, V., Happé, F., Plomin, R., & Ronald, A. (2011). Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%., and 1%). *Archives of General Psychiatry, 68*(11), 1113-1121. doi: 10.1001/archgenpsychiatry.2011.119

Robinson, E., Munir, K., Munafò, M., Hughes, M., McCormick, M., & Koenen, K. (2011). Stability of Autistic Traits in the General Population: Further Evidence for a Continuum of Impairment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(4), 376–384. doi: 10.1016/j.jaac.2011.01.005

Robinson, S., Goddard, L., Dritschel, B., Wisley, M., & Howlin, P. (2009). Executive functions in children with Autism Spectrum Disorders. *Brain and Cognition, 71*(3), 362–368. https://doi.org/10.1016/j.bandc.2009.06.007

Ronald, A., Happé, F., Price, T. S., Baron-Cohen, S., & Plomin, R. (2006). Phenotypic and Genetic Overlap Between Autistic Traits at the extremes of the general population. *Journal of the American Academy of Child & Adolescent Psychiatry, 45*(10), 1206-1214. doi: 10.1097/01.chi.0000230165.54117.41

Roth, R. M., & Gioia, G. A. (2005). *Behavior rating inventory of executive function--adult version*. Lutz, FL: Psychological Assessment Resources.

Rubenstein, E., & Chawla, D. (2018). Broader autism phenotype in parents of children with autism: A systematic review of percentage estimates. *Journal of Child and Family Studies, 27*(6), 1705-1720. doi: 10.1007/s10826-018-1026-3

Rylaarsdam, L., & Guemez-Gamboa, A. (2019). Genetic causes and modifiers of autism

spectrum disorder. *Frontiers in Cellular Neuroscience, 13,* 385-400. doi: 10.3389/fncel.2019.00385

Sasson, N. J., Faso, D. J., Parlier, M., Daniels, J. L., & Piven, J. (2014). When father doesn't know best: Selective disagreement between Self-Report and informant report of the broad autism phenotype in parents of a child with autism. *Autism Research, 7*(6), 731-739. doi: 10.1002/aur.1425

Sasson, N., Lam, K., Childress, D., Parlier, M., Daniels, J., & Piven, J. (2013). The Broad Autism Phenotype Questionnaire: Prevalence and Diagnostic Classification. *Autism Research, 6*(2), 134–143. doi: 10.1002/aur.1272

Scheeren, A. M., Koot, H. M., & Begeer, S. (2012). Social Interaction Style of Children and Adolescents with High-Functioning Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *42*(10), 2046–2055. https://doi.org/10.1007/s10803-012-1451-x

Schriber, R. A., Robins, R. W., & Soloman, M. (2014). Personality and self-insight in individuals with autism spectrum disorder. *Journal of Personality and Social Psychology, 106*(1), 112-130. doi: 10.10137/a0034950

Seidman, I., Yirmiya, N., Milshtein, S., Ebstein, R., & Levi, S. (2011). The broad autism phenotype questionnaire: Mothers versus fathers of children with an autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *42*(5), 837-846. doi: 10.1007/s10803-011-1315-9

Shi, L. J., Ou, J. J., Gong, J. B., Wang, S. H., Zhou, Y. Y., Zhu, F., Liu, X. D., Zhao, J. P., & Luo, X. R. (2015). Broad autism phenotype features of Chinese parents with autistic children and their associations with severity of social impairments in

probands. *BMC Psychiatry, 15*(1), 1-9. doi: 10.1186/s12888-015-0568-9

Stewart, M., & Austin, E. (2009). The structure of the Autism-Spectrum Quotient (AQ): Evidence from a student sample in Scotland. *Personality and Individual Differences, 47*(3), 224–228. doi: 10.1016/j.paid.2009.03.004

Sucksmith, E., Roth, I., & Hoekstra, R. A. (2011). Autistic traits below the clinical threshold: Re-examining the broader autism phenotype in the 21st century. *Neuropsychology Review, 21*(4), 360-389. doi: 10.1007/s11065-011-9183-9

Suh, J., Orinstein, A., Barton, M., Chen, C., Eigsti, I., Ramirez-Esparza, N., & Fein, D. (2016). Rating of the broader autism phenotype and personality traits in optimal outcomes from autism spectrum disorder. *Journal of Autism and Developmental Disorders, 46*(11), 3505-3518. doi: 10.1007/s10803-016-2868-4

Szatmari, P., MacLean, J., Jones, M., Bryson, S., Zwaigenbaum, L., Bartolucci, G., …Tuff, L. (2000). The familial aggregation of the lesser variant in biological and nonbiological relatives of PDD probands: A family history study. *Journal of Child Psychology and Psychiatry, 41*(5), 579-586. doi: 10.1111/14697610.00644

Thapar, A., & Cooper, M. (2013). Copy number variation: What is it and what has it told us about child psychiatric disorders?. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(8), 772-774. doi: 10.1016/j.jaac.2013.05.013

Toplak, M. E., West, R. F., Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and rating of executive function assess the same construct?. *The Journal of Child Psychology and Psychiatry, 54*(2), 131-143. doi: 10.1111/jcpp.12001

Van der Elst, W., Ouwehand, C., van der Werf, G., Kuyper, H., Lee, N., & Jolles, J. (2012). The Amsterdam Executive Function Inventory (AEFI): Psychometric properties and

demographically corrected normative data for adolescents aged between 15 and 18 years. *Journal of Clinical and Experimental Neuropsychology, 34*(2), 160–171. doi: 10.1080/13803395.2011.625353

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486–492. doi: 10.1080/17405629.2012.686740

Vickerstaff, S., Heriot, S., Wong, M., Lopes, A., & Dossetor, D. (2007). Intellectual ability, self-perceived social competence, and depressive symptomology in children with high-functioning autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 37,* 1647-1664. doi: 10.1007/s10803-006-0292-x

Wainer, A. L., Ingersoll, B. R., & Hopwood, C. J. (2011). The structure and nature of the broader autism phenotype in a non-clinical sample. *Journal of Psychopathology and Behavioural Assessment, 33*(4), 459-469. doi: 10.1007/s10862-011-9259-0

Wakabayashi, A., Baron-Cohen, S., & Wheelwright, S. (2006). Are autistic traits an independent personality dimension? A study of the Autism-Spectrum Quotient (AQ) and the NEO-PI-R. *Personality and Individual Differences, 41*(5), 873–883. doi: 10.1016/j.paid.2006.04.003

Warren, Z., Foss-Feig, J., Malesa, E., Lee, E., Taylor, J., Newsom, C., Crittendon, J., & Stone, W. (2012). Neurocognitive and Behavioral Outcomes of Younger Siblings of Children with Autism Spectrum Disorder at Age Five. *Journal of Autism and Developmental Disorders, 42*(3), 409–418. doi: 10.1007/s10803-011-1263-4

Wheelwright, S., Auyeung, B., & Allison, C. (2010). Research defining the broader, medium, and narrow autism phenotype among parents using the Autism Spectrum Quotient

(AQ). *Molecular Genetics, 1,* 10-19. doi: 10.1186/2040-2392-1-10

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. *Journal of Autism and Developmental Disorders, 9*(1), 11–29. https://doi.org/10.1007/BF01531288

Xiao, L., & Hau, K.-T. (2023). Performance of Coefficient Alpha and Its Alternatives: Effects of Different Types of Non-Normality. *Educational and Psychological Measurement, 83*(1), 5–27. https://doi.org/10.1177/00131644221088240

Table 1

Sociodemographic characteristics (N = 899)

Characteristic	Full Sample	
	n	%
Biological Sex		
Female	697	77.4
Male	196	21.8
Prefer not to say	2	0.2
Gender identity		
Female	685	76.1
Male	195	21.7
Transgender	3	0.3
Non-Binary	9	1.0
Gender Fluid	1	0.1
Other	3	0.3
Ethnicity		
Arab/West Asian	12	1.3
Black/African Canadian	109	12.1
Hispanic/Latino	2	0.2
Chinese	20	2.2
Filipino	74	8.2
Indigenous	59	6.6
Japanese	1	0.1
Korean	2	0.2
Latin American	3	0.3
South Asian	45	5.0
South East Asian	12	1.3
White/Caucasian	514	57.1
Other	34	3.8
Prefer not to say	9	1.0
Relationship status		
Single	491	54.6
Dating	280	31.1
Living together	35	3.9
Engaged	8	0.9
Married	53	5.9
Common-Law	19	2.1
Separated	3	0.3
Divorced	1	0.1
Prefer not to say	6	0.7
Sexual Orientation		
Heterosexual	702	78.0
Gay	9	1.0
Lesbian	9	1.0
Bisexual	93	10.3

Pansexual	25	2.8
Asexual	8	0.9
Other	25	2.8
Prefer not to say	24	2.7
Highest education level achieved		
High school	501	55.7
Some college	51	5.7
College degree	22	2.4
Some university	273	30.3
Associates/Bachelor's degree	35	3.9
Master's degree	>5	0.3
Other	10	1.1
Prefer not to say	1	0.1
University program		
Faculty of Arts	145	16.1
Faculty of Science	227	25.2
University 1	392	43.6
Faculty of Agricultural & Food Sciences	12	1.3
School of Art	2	0.2
School of Business	40	4.4
Faculty of Education	3	0.3
Faculty of Engineering	1	0.1
Faculty of Environment, Earth, & Resources	2	0.2
Extended Education	19	2.1
Faculty of Health Sciences	10	1.1
College of Nursing	1	0.1
School of Dental Hygiene	1	0.1
Kinesiology and Recreation Management	28	3.1
Faculty of Music	3	0.3
Faculty of Social Work	3	0.3
Other	4	0.4
Prefer not to say	3	0.3

Note. N = 899. Participants were on average 21.3 years old (SD = 11.70).

Table 2

Mental Health Diagnoses

Diagnosis	*n*	%
Generalized Anxiety Disorder	159	17.7
Social Anxiety Disorder	82	9.1
Panic Disorder	19	2.1
Obsessive Compulsive Disorder	22	2.4
Posttraumatic Stress Disorder	28	3.1
Major Depressive Disorder	63	7.0
Persistent Depressive Disorder	9	1.0
Bipolar I Disorder	4	0.4
Bipolar II Disorder	3	0.3
Cyclothymia	1	0.1
Other mental health diagnosis	47	5.2
No mental health diagnosis	640	71.1

Note. Numbers reflect the number of participants who reported “yes” in response to questions asking if they have been diagnosed with a particular disorder.

Table 3

Family History of ASD

Response	*n*	%
Yes	87	9.7
No	724	80.4
Unsure	80	8.9
Prefer not to say	1	0.1

Table 4

Means and Standard Deviations for Measures

Variable	Total sample *M*(*SD*)	Female *M*(*SD*)	Male *M*(*SD*)
Broad Autism Phenotype Questionnaire			
Total score	2.99(.57)	3.00(.59)	2.95(.52)
Aloof	2.92(.78)	2.91(.79)	2.95(.73)
Rigid	3.26(.71)	3.28(.72)	3.16(.64)
Pragmatic Language	2.61(.66)	2.63(.67)	2.53(.62)
Big Five Personality Inventory			
Openness	3.48(.54)	3.48(.55)	3.49(.50)
Conscientiousness	3.59(.67)	3.59(.68)	3.58(.64)
Extraversion	3.05(.80)	3.07(.78)	2.97(.87)
Agreeableness	3.91(.60)	3.93(.60)	3.85(.57)
Neuroticism	3.36(.79)	3.48(.77)	2.93(.69)
CC-SR Pragmatic Language	1.45(.38)	1.45(.40)	1.45(.34)
Amsterdam Executive Functioning Inventory			
Total score	1.92(.31)	1.94(.32)	1.84(.27)
Attention	2.15(.58)	2.17(.58)	2.10(.55)
Planning	2.06(.31)	2.10(.30)	1.98(.33)
Self-Control/Self-Monitoring	1.63(.50)	1.66(.52)	1.53(.45)

Note. CC-SR = Communication Checklist Self-Report.

Table 5

Goodness-of-Fit Indicators for the Original Three-Factor Model of the BAPQ

Goodness-of-Fit Indicators	χ^2	*df*	RMSEA	CFI	TLI	SRMR
	3476.64*	591	.07	.73	.71	.08

* $p < .00001$

Table 6

Model Assessment of Configural Invariance

Model	Chi-square	Chi-square df	CFI	TLI	RMSEA	SRMR
Ungrouped unconstrained model	3476.640	591	0.731	0.713	0.074	0.076
Grouped unconstrained model	4262.459	1182	0.719	0.701	0.077	0.081
Difference	785.819	591	0.012			

Table 7

Model Assessment of Metric Invariance

Model	Chi-square	Chi-square df	CFI	TLI	RMSEA	SRMR
Unconstrained model	4262.459	1182	0.719	0.701	0.077	0.081
Factor loadings constrained model	4297.329	1215	0.719	0.709	0.076	0.082
Difference	34.87	33	0			

Table 8

Model Assessment of Structural Invariance

Model	Chi-square	Chi-square df	CFI	TLI	RMSEA	SRMR
Factor loadings constrained model	4297.329	1215	0.719	0.709	0.076	0.082
Factor covariance constrained model	4490.166	1248	0.705	0.702	0.076	0.085
Difference	192.837	33	0.014			

Table 9

Confirmatory Factor Analysis: Standardized Factor Loadings for the Original Model of the BAPQ

Subscale	Items	Factor loadings
Aloof	(1) I like being around other people	.75
	(5) I would rather talk to people to get information than to socialize	.46
	(9) I enjoy being in social situations	.81
	(12) People find it easy to approach me	.52
	(16) I look forward to situations where I can meet new people	.73
	(18) When I make conversation it is just to be polite	.40
	(23) I am good at making small talk	.60
	(25) I feel like I am really connecting with other people	.68
	(27) Conversation bores me	.52
	(28) I am warm and friendly in my interactions with others	.46
	(31) I prefer to be alone rather than with others	.65
	(36) I enjoy chatting with people	.83
Pragmatic Language	(2) I find it hard to get my words out smoothly	.60
	(4) It's hard for me to avoid getting sidetracked in conversation	.43
	(7) I am "in tune" with the other person during conversation	.60
	(10) My voice has a flat or monotone sound to it	.40
	(11) I feel disconnected or "out of sync" in conversations with others	.71
	(14) People ask me to repeat things I've said because they don't understand me	.55
	(17) I have been told that I talk too much about certain topics	.28
	(20) I speak too loudly or softly	.44
	(21) I can tell when someone is not interested in what I am saying	.10
	(29) I leave long pauses in conversation	.47
	(32) I lose track of my original point when talking to people	.58
	(34) I can tell when it is time to change topics in conversation	.42
Rigid	(3) I am comfortable with unexpected changes in plans	.56
	(6) People have to talk me into trying something new	.55
	(8) I have to warm myself up to the idea of visiting an unfamiliar place	.59
	(13) I feel a strong need for sameness from day to day	.58
	(15) I am flexible about how things should be done	.53
	(19) I look forward to trying new things	.57
	(22) I have a hard time dealing with changes in my routine	.65
	(24) I act very set in my ways	.42
	(26) People get frustrated by my unwillingness to bend	.46
	(30) I alter my daily routine by trying something different	.41
	(33) I like to closely follow a routine while working	.35
	(35) I keep doing things the way I know, even if another way might be better	.50

Table 10

Goodness-of-Fit Indicators for Model 2

Goodness-of-Fit Indicators	χ^2	*df*	RMSEA	CFI	TLI	SRMR
	2490.56*	519	.06	.81	.79	.07

* $p < .00001$

Table 11

Correlations Between Total Scale and Subscale Sumscores for Total Sample

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. BAPQ total	1.00													
2. CCSR - Pragmatic Lang total	.49**	1.00												
3. Rigid	.78**	.35**	1.00											
4. Pragmatic Lang	.76**	.62**	.39**	1.00										
5. Aloof	.84**	.24**	.50**	.48**	1.00									
6. Extraversion	-.53**	-.02	-.32**	-.26**	-.67**	1.00								
7. Agreeableness	-.42**	-.24**	-.29**	-.35**	-.39**	.14**	1.00							
8. Conscientious	-.40**	-.35**	-.16**	-.52**	-.31**	.16**	.41**	1.00						
9. Neuroticism	.50**	.32**	.43**	.44**	.35**	-.25**	-.35**	-.39**	1.00					
10. Openness	-.19**	.03	-.23**	-.04	-.16**	.19**	.14**	.08*	-.09**	1.00				
11. AEFI Total	.34**	.49**	.26**	.53**	.08*	.23**	-.29**	-.44**	.38**	.03	1.00			
12. Attention	.43**	.40**	.26**	.55**	.26**	-.10**	-.26**	-.60**	.42**	.01	.74**	1.00		
13. Planning	-.11**	-.05	.009	-.14**	-.12**	.13**	.10**	.41**	-.07**	.08*	.16**	-.19**	1.00	
14. Self-Control	.21**	.44**	.18**	.41**	-.05	.38**	-.27**	-.35**	.27**	-.001	.84**	.37**	-.05	1.00

Note. BAPQ total = BAPQ total score; CCSR – Pragmatic Lang total = CCSR Pragmatic Language subscale score; Pragmatic Lang = Pragmatic Language subscale score; Conscientious = Conscientiousness subscale score; AEFI = Amsterdam Executive Functioning total score.

* indicates <.05 (two-tailed); ** indicates p <.01 (two-tailed).

Table 12

Correlations Between Total Scale and Subscale Sumscores for Males

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. BAPQ total	1.00													
2. CCSR - Pragmatic Lang total	.40**	1.00												
3. Rigid	.73**	.30**	1.00											
4. Pragmatic Lang	.76**	.54**	.34**	1.00										
5. Aloof	.82**	.14	.40**	.44**	1.00									
6. Extraversion	-.56**	-.04	-.27**	-.33**	-.68**	1.00								
7. Agreeableness	-.42**	-.21**	-.33**	-.30**	-.35**	.14**	1.00							
8. Conscientious	-.35**	-.30**	-.22**	-.45**	-.20**	.16*	.38**	1.00						
9. Neuroticism	.44**	.27**	.35**	.40**	.31**	-.33**	-.30**	-.36**	1.00					
10. Openness	-.16*	.04	-.29**	.01	-.08	.25**	.11	.07	-.05	1.00				
11. AEFI Total	.23**	.42**	.29**	.40**	-.04	.24**	-.34**	-.42**	.26**	.07	1.00			
12. Attention	.40**	.30**	.26**	.46**	.20**	-.17**	-.25**	-.62**	.40**	.03	.67**	1.00		
13. Planning	-.11**	-.04	.05	-.16*	-.10	.17*	.11	.45**	-.15*	.06	.13	-.32**	1.00	
14. Self-Control	.07	.38**	.18*	.27**	-.18*	.42**	-.34**	-.30**	.12	.03	.81**	.27**	-.06	1.00

Note. BAPQ total = BAPQ total score; CCSR – Pragmatic Lang total = CCSR Pragmatic Language subscale score; Pragmatic Lang = Pragmatic Language subscale score; Conscientious = Conscientiousness subscale score; AEFI = Amsterdam Executive Functioning total score.

* indicates <.05 (two-tailed); ** indicates p <.01 (two-tailed).

Table 13

Correlations Between Total Scale and Subscale Sumscores for Females

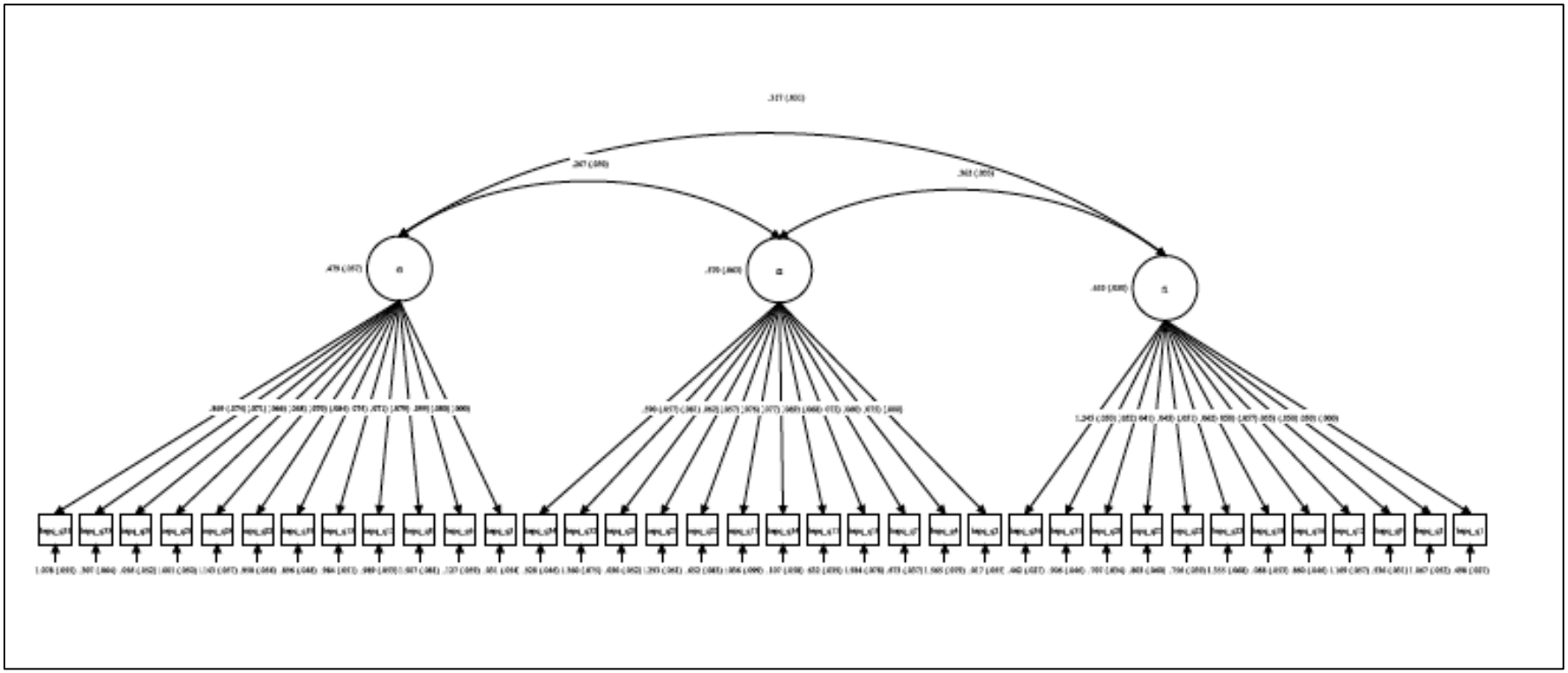
Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. BAPQ total	1.00													
2. CCSR - Pragmatic Lang total	.49**	1.00												
3. Rigid	.79**	.36**	1.00											
4. Pragmatic Lang	.76**	.63**	.39**	1.00										
5. Aloof	.85**	.25**	.53**	.49**	1.00									
6. Extraversion	-.53**	-.01	-.34**	-.24**	-.67**	1.00								
7. Agreeableness	-.43**	-.24**	-.28**	-.37**	-.40**	.14**	1.00							
8. Conscientious	-.41**	-.36**	-.14**	-.54**	-.33**	.16**	.42**	1.00						
9. Neuroticism	.53**	.36**	.45**	.45**	.38**	-.27**	-.40**	-.42**	1.00					
10. Openness	-.20**	.01	-.23**	-.06	-.18**	.17**	.15**	.09*	-.11**	1.00				
11. AEFI Total	.36**	.51**	.25**	.55**	.10*	.23**	-.29**	-.45**	.38**	.02	1.00			
12. Attention	.44**	.42**	.25**	.57**	.27**	-.08*	-.26**	-.60**	.43**	.004	.75**	1.00		
13. Planning	-.12**	-.05	-.01	-.15**	-.12**	.12**	.10*	.40**	-.11**	.08*	.15**	-.17**	1.00	
14. Self-Control	.23**	.45**	.17**	.44**	-.02	.37**	-.26**	-.36**	.28**	-.01	.84**	.39**	-.06	1.00

Note. BAPQ total = BAPQ total score; CCSR – Pragmatic Lang total = CCSR Pragmatic Language subscale score; Pragmatic Lang = Pragmatic Language subscale score; Conscientious = Conscientiousness subscale score; AEFI = Amsterdam Executive Functioning total score.

* indicates <.05 (two-tailed); ** indicates p <.01 (two-tailed).

Figure 1

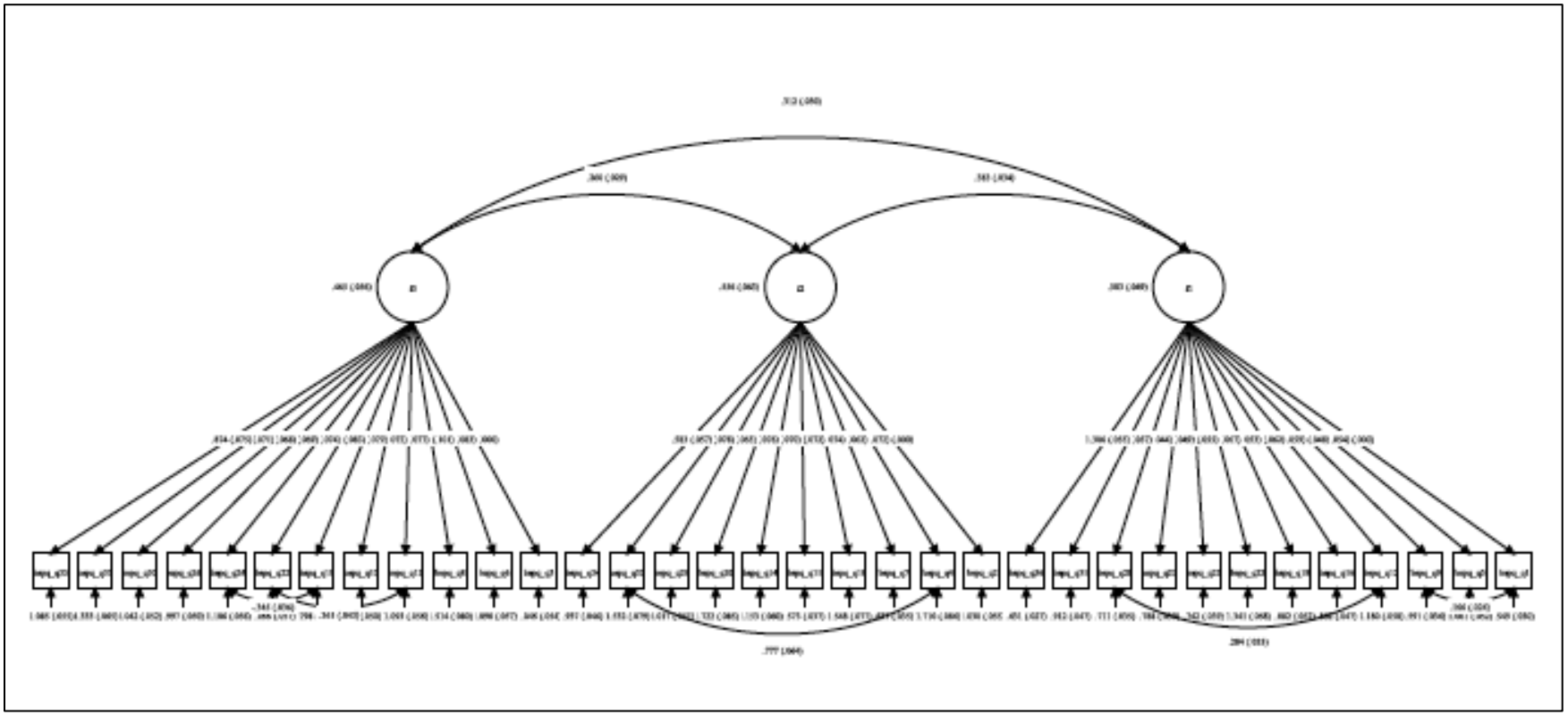
Path Diagram of the BAPQ Original Model



Note. BAPQ = Broad Autism Phenotype Questionnaire.

Figure 2

Path Diagram of Model 2 of the BAPQ



Note. BAPQ = Broad Autism Phenotype Questionnaire.

Appendix

(On Department letterhead)
Letter of Informed Consent

Information and Consent Form

Study Name: The Factor Structure of the Broad Autism Phenotype Questionnaire in a Non-Clinical Sample of Canadian Undergraduate Students

Principal Investigator: Sydney Kingston, PhD Candidate, Psychology, _____

Research Supervisor: Dr. Janine Newton Montgomery, Associate Professor, Psychology, _____

This consent form, a copy of which will be left with you for your records and reference, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, you should feel free to ask. Please take the time to read this carefully and to understand any accompanying information.

Purpose of Study

Sydney Kingston is conducting this study as her PhD dissertation, under the supervision of Dr. Janine Newton Montgomery. The purpose of this research is to examine the factor structure and construct validity of the Broad Autism Phenotype Questionnaire in a sample of Canadian undergraduate students. Undergraduate students in Introductory Psychology (PSYC 1200) will be recruited to take part in this study for course credit. Participants will first complete a brief screening questionnaire to determine study eligibility. Eligible participants will then complete an online questionnaire which include demographic questions such as age, biological sex, gender identity, ethnicity, and physical and mental health diagnoses. This is followed by several questionnaires which assess self-perceived personality characteristics, social abilities, interests, executive functioning, and pragmatic language abilities. We estimate that the time needed to complete this study will vary from participant to participant; however, our estimated time to completion is 30 minutes to one hour. Participants will earn two course credits for their participation.

Risks and Discomforts

There is a small risk that some participants may feel uncomfortable answering some of these questions. You may choose to discontinue your participation at any time and/or choose not to answer particular questions. If you choose to participate and experience distress as a result of answering these questions, please review the list of helpful resources included at the end of the questionnaire. If answering these questions causes significant or persistent distress, please contact your local crisis line or healthcare professional.

Benefits

By participating in this study, there may not be direct benefits to you. However, an indirect benefit is that by participating in this study, you will be providing information that will help to our increase knowledge of the BAP and the psychometric properties of the Broad Autism Phenotype Questionnaire; possibly informing if and where improvements to this measure need to be made. This will help to increase Canadian researchers and clinicians' confidence that they are using a sound measure to assess BAP traits. As per PSYC 1200, you will earn course credit for your participation.

Confidentiality

Participation in this study is voluntary. All the responses you provide will be kept confidential. The information that you provide will be stored on Qualtrics which is encrypted and password protected. Once this information is taken off Qualtrics, it will be saved in a password protected database and stored on a password protected computer. Your research data will be identified by a number. This will include demographic information and questionnaire responses. Only the primary researcher, her advisor, and authorized research team members (e.g., research assistants) will have access to the data. Once data collection and analysis are complete, we will share the results of this research in the primary researcher's PhD dissertation, presentations, conferences, journal articles, and the Social Cognition Lab website. Data will only be presented in aggregate form. The data gathered for this research project will be kept indefinitely for the purposes of conducting possible future research projects by members of the Social Cognition Lab. The data will be stored in a password protected database on a password protected computer. After the data is downloaded, it will be deleted off Qualtrics.

Online research has an inherently higher risk of confidentiality breaches than in-person research; however, Qualtrics is encrypted and approved for research by University of Manitoba.

Findings

The results of this study should be available by December 2023. You will be able to see the research summary posted on the Social Cognition Lab website once it is available (see _____. If you would like to receive a summary, please contact Sydney Kingston at _____, provide her with your University of Manitoba email address and inform her that you would like to receive a copy of the summary via email. You only need to provide this information if you wish to receive a summary of the results; you are not required to provide this information.

Voluntary Participation/Withdrawal of Participation

Your signature on this from indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a subject. In no way does this way your legal rights nor release the researchers, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from the study at any time, and/or refrain from answering questions you prefer to omit, without prejudice

or consequence. Should you feel uncomfortable and wish to withdraw, you may exit out of the survey.

If you decide to withdraw and have already submitted answers to some questions, the primary researcher is interested in using your partial data. Please indicate below, by selecting yes or no, whether you consent to having your partial responses used should you decide to withdraw.

If I decide to withdraw my participation, I consent to having my partial data used:

**in Qualtrics participants will see two options to click on <Yes> and <No>

The university of Manitoba may look at your research records to see that the research is being done in a safe and proper way.

This research has been approved by the Research Ethics Board at the University of Manitoba, Fort Garry campus. If you have any concerns or complaints about this project you may contact any of the above-named persons or the Human Ethics Officer at 204-474-7122 or HumanEthics@umanitoba.ca. A copy of this consent form has been given to you to keep for your record and reference.

Clicking “I agree” at the bottom of this page indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a subject. Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation.

If you have read the information presented in this form and do not have any questions about this study, please click “I agree” when you are ready to begin. You should only click “I agree” if you agree to participate with full knowledge of the study presented to you in this information and consent form and of your own free will. We strongly encourage you to save or print a copy of this consent form now for your own records, as it will not be available later.

Thank you for participating in this study.

**in Qualtrics participants will see two options to click on <I agree> (proceed to study), and <I do not agree> (exit)