INTER-RATER AGREEMENT IN AUTISM SPECTRUM DISORDER FOR ANXIETY, DEPRESSION, AND BROAD INTERNALIZING SYMPTOMS: A META-ANALYSIS

By

Janelle Youngdahl

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

School Psychology – Doctor of Philosophy

2020

ABSTRACT

INTER-RATER AGREEMENT IN AUTISM SPECTRUM DISORDER FOR ANXIETY, DEPRESSION, AND BROAD INTERNALIZING SYMPTOMS: A META-ANALYSIS

By

Janelle Youngdahl

Youth with Autism Spectrum Disorder (ASD) are at elevated risk for internalizing symptoms such as anxiety and depression (Bellini, 2004; Kim et al., 2000; Matson & Williams, 2014). These internalizing problems can affect self-esteem, social competence, academic performance, and physical health; thus, it is critical to accurately identify internalizing symptoms in order to provide appropriate intervention to those in need (Michael & Merrell, 1998). One of the most common ways to screen for internalizing symptoms is through use of rating scales completed by youth, parent, and/or teacher informants. However, inconsistent inter-rater agreement findings across studies of youth with ASD have rendered the literature difficult to summarize and in need of more systematic investigation. No prior meta-analysis has examined inter-rater or cross-informant ratings agreement concerning different internalizing constructs in youth with ASD specifically—despite its relevance to a multi-method and multi-informant approach to assessment typically recommended as best practice (Taylor et al., 2018). The present meta-analysis (a) closely examined the level of agreement across different rater-pairs (i.e., parent vs. youth, teacher vs. youth, and parent vs. teacher) assessing internalizing problems (i.e., anxiety, depression, and broad internalizing) in youth with ASD, (b) investigated both inter-rater correlations and cross-rater mean differences, (c) assessed potential moderator variables that could impact the magnitude or direction of correlations or mean differences, and (d) systematically summarized findings and trends.

Results indicated that across the three constructs (i.e., anxiety, depression, and broad internalizing), the mean *r* ranged from 0.399 to 0.430 (moderate range) for parent vs. youth self-report ratings and 0.256 to 0.296 (small range) for parent vs. teacher ratings. In the case of teacher vs. parent ratings, the observed mean inter-rater correlations ranged from 0.229 to 0.342 (small to moderate range) but were non-significant for all three constructs. Moderator analyses within the parent vs. youth self-report inter-rater correlations indicated that method of youth self-report administration moderated correlations for anxiety, while mean age of the youth moderated correlations for depression. No significant moderators were noted for other inter-rater correlations across the three rater-pairs. For parent vs. youth self-report standardized mean differences, mean effect size *g* was 0.220 for anxiety, 0.788 for depression, and 0.090 for broad internalizing. However, evidence of possible publication bias and associated re-estimation yielded non-significant bias-adjusted mean *g* estimates in the small to negligible range for both constructs. For parent vs. teacher ratings, mean *g* values ranged from 0.015 to 0.176, but all were deemed negligible. In the case of teacher vs. youth self-report ratings, mean *g* varied considerably, ranging from -0.033 to 0.670—but all mean *g* values were non-significant and based on only a small number of studies. No significant moderators were found for any of the standardized mean differences across all rater-pairs and constructs.

These results suggest that covariation across informants regarding internalizing symptoms in youth with ASD tends to be small to moderate, depending on the rater-pairs, and typically involves negligible mean differences between rater types. Additional inter-rater studies are needed, in general, to improve precision of effect size estimates and provide additional power for moderator analyses, but are needed, in particular, for teacher vs. youth self-report ratings—where overall estimates are based on too few studies.

Neil
Mom and Dad
Jaclyn, Zach, and Isabella

ACKNOWLEDGEMENTS

I would like to first thank my family for their unwavering love and support. To my husband, Neil—thank you for your calming presence, positivity, and reassurance. You mean the world to me. To my parents, John and Jeannine—thank you does not begin to express my gratitude for you. I am lucky to have you two as role models. I carry your unconditional love with me always. To my sister, Jaclyn—thank you for your fierce love and support throughout my life. I know I always have you in my corner. To my best friend and Ph.D. cohort mate, Danielle—thank you for your humor, advice, and encouragement. I could not have hand-picked a better person to be by my side for the past five years.

I would also like to thank the numerous people in the MSU School Psychology program who have helped me along the way. To my advisor, Dr. Martin Volker—thank you for sharing your expertise and wisdom with me over the past five years. I am grateful for your guidance, support, and dedication to thoroughness. To my dissertation committee, Dr. Jodene Fine, Dr. Marisa Fisher, and Dr. Gloria Lee—thank you for your encouragement and dedication to providing me with valuable feedback. To the entire MSU School Psychology faculty—thank you for your leadership and commitment to students. I am forever thankful for the education and training I received in our program. To Dr. Jana Aupperlee—I had the pleasure of being taught by you, alongside my small cohort, for many of my graduate school courses. Thank you for your support, example of professionalism, and positivity over the years. To Nicole, Ellen, and Megan—thank you for your enthusiastic help with searching for and coding articles to include in this meta-analysis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The prevalence estimates of autism spectrum disorder (ASD) continue to increase over time, with a current prevalence of one in 59 children (Baio et al., 2018). ASD is a neurodevelopmental disorder characterized by deficits in social communication and interaction, and restricted, repetitive patterns of behavior, interests, or activities (American Psychiatric Association [APA], 2013). In addition to these core features of the disorder, a variety of associated features and comorbidities may also be present. For example, intellectual disability (ID; Baio et al., 2018), language impairments (APA, 2013; Mazzone, Ruta, & Reale, 2012; Tager-Flusberg & Kasari, 2013), motor deficits (MacDonald, Lord, & Ulrich, 2014; Mazzone, Ruta, & Reale, 2012; McPhillips et al., 2014; APA, 2013), seizures (Theoharides & Zhang, 2011), attention-deficit/hyperactivity disorder (ADHD; Jang et al., 2013), and oppositional defiant disorder (ODD; Gadow et al., 2004) are observed to be more common in youth with ASD. Most essential to the present meta-analysis are the psychiatric comorbidities that include internalizing problems, such as anxiety and depression. Youth with ASD are at greater risk of experiencing anxiety and depression than the general population (Bellini, 2004; Kim et al., 2011; Matson & Williams, 2014), and these internalizing issues can have significant negative effects on the academic, social, and physical development of children and adolescents (Michael & Merrell, 1998). Additionally, the presence of anxiety or depression can exacerbate the core symptoms of ASD (Davidsson et al., 2017). Therefore, it is critically important to identify these issues in youth with ASD so that appropriate intervention and support can be initiated.

Internalizing problems can be defined using a variety of constructs, often based on which particular instrument is used to assess for such issues. However, anxiety and depression are

typically the most commonly recognized internalizing problems (Merrell, 2008). There are many ways to screen for and assess internalizing problems in youth. Specifically, clinical interviews, direct observation, and rating scales are commonly employed methods (Gray et al., 2009; Klein, Dougherty, & Olino, 2005). Clinical interviews can provide rich and detailed information about depressive and anxious symptoms; however, they can be time consuming, are not norm-referenced, and may not provide standardized measurement of the behavior (depending on the structure of the interview). Direct observation provides the unique opportunity for the evaluator to observe the child/adolescent's behavior without having to rely on reports from the child or a third-party individual (e.g., parent, teacher, medical doctor, etc.), yet it also comes with challenges related to reliability and validity—and may not be the best strategy for detecting internalizing issues. Rating scales are commonly preferred methods of assessment because they are efficient, can be completed by multiple informants, and provide ratings of behavior in a standardized (and, in some cases, norm-referenced) manner (Whitcomb & Merrell, 2013).

Although rating scales are an efficient and standardized way to assess behavioral problems, they can involve several issues with reliability and validity. That is, the level of agreement among raters (e.g., parents, children/adolescents, and teachers) can be weak, which makes it difficult to accurately identify co-occurring issues and make subsequent clinical decisions (McDonald et al., 2016). Many factors can lead to higher levels of disagreement among multiple raters such as general measurement issues, natural variation in child/adolescent behavior across settings, and variability in capacity of youth raters to accurately self-report (Humrichouse et al., 2007; Whitcomb & Merrell, 2013). Additionally, the Attribution Bias Context **(**ABC) Model (De Los Reyes & Kazdin; 2005) suggests that informant discrepancies are related to different motivations for entering the treatment process that may come with varying

perspectives on child behavior and the severity of behavior. Further, youth with ASD are more likely to experience issues such as poor self-awareness, difficulty understanding emotions, limited communication skills, presence of alexithymia, under-developed theory of mind, and/or deficits in executive functioning, that may make it difficult for them to report on their internalizing symptoms (see Baron-Cohen, 2002; Bird & Cook, 2013; Hagopian & Jennett, 2014; Kiep & Spek, 2016; Lopata et al., 2010; Mazefsky et al., 2011; Spek, Scholte, & Van Berckelaer-Onnes, 2009).

There is evidence to support potential moderators of the relationship between raters of internalizing problems. These moderators include age of the youth (Achenbach et al., 1987; Ebesutani et al., 2011; Stratis & Lecavalier, 2015;Vasa et al., 2016) cognitive abilities of the youth (Durbin, 2010; Stratis & Lecavalier, 2015), and method of self-report administration (Bitsika et al., 2015; Farrugia & Hudson, 2006; Jepsen, Gray, & Taffe, 2012; Magiati et al., 2014). Thus, the current meta-analysis will analyze these factors as moderator variables. Other moderators are also possible (e.g., score type, gender, socioeconomic status, etc.). However, they are less likely to either be reported or to provide sufficient variability across studies to properly assess.

Given the frequency with which rating scales are used to assess internalizing problems, the prevalence of anxiety and depression in youth with ASD, and the negative consequences of untreated internalizing issues for other areas of functioning, it is important to better understand the agreement among multiple raters and what variables, if any, lead to higher or poorer levels of agreement. Prior meta-analyses have systematically reviewed related topics such as cross-informant agreement of emotional/behavioral issues in diverse or typically developing (TD) youth samples (Achenbach et al., 1987; Huang, 2007) or adult populations (Achenbach,

Krukowski, Dumenci, and Ivonova, 2005; Hollocks et al., 2018), prevalence rates of anxiety and depression within the ASD population (Hudson, Hall, & Harkness, 2018; Van Steensel & Heeman, 2017), and cross-informant agreement of broad emotional/behavioral issues within ASD and ID (without ASD) populations (Stratis & Lecavalier, 2015). However, these meta-analyses do not cover multi-informant agreement within youth with ASD specifically, both correlational and mean difference estimates of agreement/disagreement within ASD, nor specific domains of internalizing beyond the broader internalizing construct in the inter-rater context.

To date, individual studies investigating this topic have found mixed results with varying levels of congruence among raters; the number of studies has increased significantly over the past decade; and there is currently not a published meta-analysis summarizing these findings that could potentially clarify the conditions that may affect agreement or lack thereof. Therefore, there is a clear need to synthesize the available information and provide insight on the patterns of multi-informant agreement in this population. Such an analysis could bring meaningful structure to the diverse and confusing array of findings, potentially clarify the conditions under which different results occur, inform clinical practice regarding what the current state of the literature would support or suggest, and reveal clear areas of need for future research. The present meta-analysis will (a) closely examine the level of agreement across different combinations of rater-pairs (i.e., parent vs. youth, teacher vs. youth, and parent vs. teacher) assessing internalizing problems (i.e., anxiety, depression, and broad internalizing) in youth with ASD, (b) investigate both inter-rater correlations and cross-rater mean differences, (c) assess potential moderator variables (e.g., youth age, youth cognitive ability, and method of self-report administration) that could affect the magnitude or direction of correlations or mean differences, and (d) systematically summarize findings and trends.

CHAPTER II

LITERATURE REVIEW

Internalizing symptoms, such as anxiety and depression, occur at higher rates in youth with ASD compared to the general population, which is problematic given the negative impact that these symptoms can have on children and adolescents (Bellini, 2004; Kim et al., 2011; Matson & Williams, 2014; Michael & Merrell, 1998). It is important to accurately identify internalizing symptoms in youth with ASD in order to determine appropriate support. This identification can be difficult given the varying levels of agreement among multiple informants who are each rating internalizing symptoms in one individual using behavior rating scales (McDonald et al., 2016).

The following sections will provide a review of the literature that is relevant to understanding cross-informant ratings of internalizing symptoms in youth with ASD. Specifically, the following sections include: (a) the definition and prevalence of ASD, (b) associated features and comorbid conditions of ASD, (c) internalizing problems and how they are assessed, (d) method and measurement issues in cross-informant agreement, (e) relevant theories related to cross-informant agreement in ASD, (f) an overview of prior meta-analyses, (g) potential moderator variables that may influence agreement across multiple informants, (h) the need for the current meta-analysis, and (i) research questions and hypotheses.

Autism Spectrum Disorder

ASD is a neurodevelopmental disorder that involves persistent deficits in social communication and social interaction across multiple contexts as manifested by deficits in social-emotional reciprocity, nonverbal communicative behaviors, and developing, maintaining, and understanding relationships. Additionally, ASD includes the presence of restricted,

repetitive patterns of behavior, interests, or activities related to stereotypies, ritualized patterns, inflexible adherence to routine, restricted interests, and/or sensory sensitivity or unusual sensory interests (APA, 2013). Current estimates indicate that the prevalence of ASD is one in 59 children (Baio et al., 2018), which is an increase from the 2014 estimate of one in 68 children (Christensen et al., 2016).

Currently, in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA, 2013), the diagnosis of ASD reflects a broadening of the autism diagnostic construct to incorporate the view of it as a "spectrum" with variability in functional levels, language involvement, and other associated features. This broadening resulted in ASD subsuming several previous related diagnoses from the fourth edition of the manual (DSM-IV-TR;APA, 2000) such as autistic disorder, Asperger's disorder, and most cases of pervasive developmental disorder not otherwise specified (PDD-NOS). Thus, individuals with a diagnosis of ASD can vary significantly in terms of level of functioning. Based on the most recent Center for Disease Control and Prevention (CDC) estimates, 69% of children identified with ASD have an IQ greater than 70, while 31% have an IQ less than or equal to 70 (Baio et al., 2018). Though definitions vary across studies, in general, high-functioning autism spectrum disorder (HFASD) is distinguishable based on language skills and cognitive ability; that is, individuals with HFASD meet the criteria for ASD, but demonstrate language and cognitive abilities no lower than the borderline range (Volker, 2012). Specifically, an IQ score of 70 is commonly considered to be the cut-off IQ score that distinguishes high- and low-functioning ASD (Lecavalier, 2014).

The prevalence rate of ASD continues to increase, from a previous estimate of 1in 150 (Rice, 2007) to the current estimate of one in 59, making it ever more critical to understand the disorder. Gaining a better understanding of ASD requires recognition and knowledge of the

common associated features of ASD and other co-morbid symptoms/disorders, as these additional features and conditions can lead to further clinical impairment and additional burden of illness on youth with ASD (Leyfer et al., 2006).

Associated Features of ASD

In addition to the core features of ASD (i.e., deficits in social communication and presence of restricted, repetitive patterns of behavior, interests, or activities), there are also a variety of associated features and comorbid conditions. That is, when compared to the general population, individuals with ASD tend to show higher rates of such issues as intellectual impairment, language impairment, motor deficits, disruptive/challenging behavior, seizures, and psychiatric conditions (APA, 2013).

To begin, intellectual disability (ID) commonly co-occurs with ASD; in fact, 31% of children identified with ASD also had comorbid ID (Baio et al., 2018). ID is characterized by substantive deficits in both general intellectual functioning and in adaptive behavior (APA, 2013); as such, comorbid ID in the context of ASD generally results in greater overall impairment and support needs (Nebel-Schwalm & Worley, 2014). Specifically, in children with ASD, IQ has been found to be a strong predictor of general functioning and a reliable indicator of prognosis (Goodwin, Matthews, & Smith, 2017).

Next, individuals with ASD may present with unusual speech patterns or language impairments such as pedantic speech (i.e., overly formal speech), monotonous or exaggerated intonation, delayed language or lack of expressive language, and/or language comprehension deficits (APA, 2013; Mazzone, Ruta, & Reale, 2012; Tager-Flusberg & Kasari, 2013). Children with ASD who do develop verbal communication tend to achieve language milestones on average 18 months later than typically developing children. In addition, children with ASD who

show delayed language acquisition have a generally poorer prognosis—particularly those who do not develop functional expressive language at all (Mayo et al., 2013).

Motor deficits are also common in individuals with ASD. For example, they may have weakened gross- and fine-motor skills that result in motor clumsiness, odd gait, poor hand-eye coordination (i.e., difficulty catching and throwing a ball), and/or walking on tiptoes (MacDonald, Lord, & Ulrich, 2014; Mazzone, Ruta, & Reale, 2012; McPhillips et al., 2014; APA, 2013). Further, many children with ASD have impairments in motor planning such as incorrect body position, poor movement timing, and increased time to initiate imitation tasks (Kaur, Srinivasan, & Bhat, 2018).

Destructive and/or self-injurious behaviors also occur at a higher rate within ASD compared to the general population. Specifically, Nebel-Schwalm and Worley (2014) indicated that symptoms associated with disruptive behavior have been found to be more frequently reported than other psychiatric symptoms in those with ASD. Common challenging and disruptive behaviors can include head banging, biting, verbal or physical aggression, and/or self-injurious behaviors (Minshawi, 2008; APA, 2013; Nebel-Schwalm & Worley, 2014).

Seizures are also common in youth with ASD; in fact, rates of occurrence can be up to ten times higher than in the general population (Theoharides & Zhang, 2011). On average, 20% to 30% of youth with ASD develop epilepsy (CDC, 2018). Children at the greatest risk for developing a seizure disorder are those who are the most severely impaired (e.g., children with both ASD and ID; Tuchman, 2013).

Finally, individuals with ASD are at-risk for psychiatric comorbidities. That is, comorbidity rates for psychological disorders and ASD range from approximately 40%-70% (Nebel-Schwalm & Worley, 2014). Some common comorbid psychological disorders include

oppositional defiant disorder (ODD), attention-deficit/hyperactivity disorder (ADHD), and internalizing disorders (e.g., anxiety and depression).

Oppositional Defiant Disorder. The prevalence of children diagnosed with ASD who also meet criteria for ODD has been estimated between 20% to 40% (Gadow et al., 2004) with comorbid disruptive behavior being a frequently endorsed psychiatric symptom in individuals with ASD (De Bruin et al., 2007). However, some of these problem behaviors, such as verbal and physical aggression, do not qualify as diagnostic of ODD in isolation (Nebel-Schwalm & Worley, 2014). Thus, those with ODD are a narrower subset of those with disruptive behaviors in the context of ASD.

Attention-Deficit/Hyperactivity Disorder (ADHD). ADHD is also frequently comorbid with ASD, with comorbidity rates ranging widely from 14% to 78% (Jang et al., 2013), and clearly a greater portion of children with ASD exhibiting symptoms of ADHD than observed in typically developing children (Nebel-Schwalm & Worley, 2014). Prior to the DSM-5, a comorbid diagnosis of ASD and ADHD was not permitted. That is, if there was already a diagnosis of one of the DSM-IV-TR pervasive developmental disorders (some of which are now subsumed under ASD in the DSM-5), a diagnosis of ADHD could not be made. This was due to the belief that symptoms of ADHD were part of the ASD disposition; however, it is now commonly believed that ASD and ADHD symptoms are distinguishable and do not always occur together in ASD (Nebel-Schwalm & Worley, 2014). It is currently hypothesized that the comorbidity of ASD and ADHD is related to deficits in executive functioning that are common in both disorders (Pitzianti et al., 2016)—suggesting possibly related neurological pathways involved in both conditions (Johnson et al., 2015). ASD and comorbid internalizing disorders

will be discussed in the following, more extensive, section given the topic's greater relevance to the present study.

Defining Internalizing Problems

Anxiety, depression, and related internalizing problems are understood to be developed and maintained within the individual. The general category of internalizing problems includes depression, anxiety, social withdrawal, and somatic/physical problems (Merrell, 2008). Merrell (2008) also indicated that anxiety, depression, and related internalizing disorders are thought to reflect over-controlled symptoms, which manifest when individuals attempt to maintain control of their internal states to an excessive degree. In contrast, externalizing problems are considered to reflect under-controlled behavior. To some degree, both dimensions are viewed as reflective of difficulties with some aspect of self-regulation.

Internalizing problems have been operationalized in several broad-band behavior rating scales. On the Child Behavior Checklist (CBCL) the Internalizing composite includes the Anxious/Depressed, Withdrawn/Depressed, and Somatic Complaints scales (Achenbach & Rescorla, 2001). On the Behavior Assessment System for Children—Third Edition (BASC-3), the Internalizing Problems composite includes the Anxiety, Depression, and Somatization scales (BASC-3; Reynolds & Kamphaus, 2004). However, anxiety and depression are the most commonly known internalizing problems (Merrell, 2008).

Anxiety. Anxiety disorders are characterized by excessive fear or anxiety and avoidant behavior (APA, 2013). Other characteristics include negative/unrealistic thoughts, panic attacks, obsessions and/or compulsive behavior, physiological arousal, and general worries (Merrell, 2008). The DSM-5 also states that anxiety disorders differ from normative fear and anxiety by being excessive or persisting beyond developmentally appropriate periods. There are many

specific disorders that fall in the category of anxiety disorders including separation anxiety disorder, selective mutism, specific phobia, social anxiety disorder (social phobia), panic disorder, agoraphobia, generalized anxiety disorder, substance/medication-induced anxiety disorder, anxiety disorder due to another medical condition, other specified anxiety disorder, and unspecified anxiety disorder (APA, 2013). Thus, it is thought that anxiety disorders may be the largest category of internalizing disorders in children (Merrell, 2008). It is estimated that 7.1% of children (ages 3-17 years) have a diagnosed anxiety disorder (CDC, 2014).

Depression. Depression is likely the most recognized and best understood internalizing disorder (Merrell, 2008). The features of depressive disorders include the presence of sad, empty, or irritable mood that significantly affects the individual's functioning; more specific characteristics can include loss of interest in activities, sleeping problems, psychomotor retardation, fatigue/lack of energy, feelings of worthlessness or guilt, and difficulty thinking/concentrating (APA, 2013). Specific disorders in the category of depressive disorders include disruptive mood dysregulation disorder, major depressive disorder, persistent depressive disorder, premenstrual dysphoric disorder, substance/medication-induced depressive disorder, depressive disorder due to another medical condition, other specified depressive disorder, and unspecified depressive disorder (APA, 2013). The prevalence of diagnosed depressive disorders in children (ages 3-17 years) is 3.2% (CDC, 2014).

Anxiety and Depression in ASD. Among the most common psychiatric comorbidities within ASD include internalizing issues such as anxiety and depression (DSM-5, 2013; Davidsson et al., 2017; Lopata et al., 2010; Park et al., 2013; Strang et al., 2012; van Steensel & Heeman, 2017). It is recognized that anxiety and depression are more common within the ASD population than in the general population (Bellini, 2004; Kim et al., 2011; Matson & Williams,

2014)—with the prevalence rate of anxiety disorders in youth with ASD ranging from 42% to 79% (Kent & Simonoff, 2017) and the prevalence rate of depression in youth with ASD ranging from 1.4% to 30% (Anderson et al., 2015). The large ranges in these prevalence rates are likely due to problems with estimating prevalence rates of anxiety and depression, because of differences in measurement or diagnostic ascertainment across studies and the difficulty of detecting internalizing issues within ASD specifically.

The presentation of anxiety in youth with ASD may include negative thoughts, obsessive-compulsive symptoms, physiological reactions, physical injury fears, specific phobias, and/or social avoidance (Kerns & Kendall, 2014). Manifestations of anxiety that are more common in individuals with ASD than in the general population can include idiosyncratic specific fears (e.g., fear of toilets, weather, etc.), increases in repetitive behaviors or intense interests, increases in sensory behaviors or sensitivity, or increases in challenging or disruptive behaviors (Magiata, Ozsivadjian, & Kerns, 2017). In individuals with HFASD, social anxiety may also be common. That is, youth with HFASD may be more prone to develop social anxiety related to experiencing increased social pressures, while having social skill deficits that make it more difficult to navigate social relationships and/or have successful social interactions (Kerns & Kendall, 2014). Additionally, social anxiety may be related to confusion among those with ASD/HFASD about social "rules" and etiquette and how to translate them into expected behavior (Magiata, Ozsivadjian, & Kerns, 2017).

Depression in youth with ASD presents similarly to depression in youth without ASD and can include tantrums, anger outbursts, fatigue, irritability, and/or loss of appetite (Ghaziuddin, Ghaziuddin, & Greden, 2002); however, in addition, ASD symptoms (e.g., poor eye contact and perseveration) may be exacerbated by comorbid depression (Ozinci, Kahn, &

Antar, 2012). Children and adolescents with HFASD are thought to be at higher risk for depression than typically developing youth. This could be due to multiple factors involving characteristics of ASD. For example, individuals with ASD tend to have more negative social experiences and/or social failures than their typically developing peers, which can be particularly damaging for youth who are higher-functioning and who may be more interested in obtaining and maintaining social relationships (Lopata et al., 2010). Also, youth with HFASD may be more attuned to recognizing their differences and may become more discouraged by them as a result (Ozinci, Kahn, & Antar, 2012).

Anxiety and depression can have a substantial impact on a youth's self-esteem, physical health, and social competence, which is why early and accurate identification of internalizing symptoms in children is important (Michael & Merrell, 1998). Additionally, these internalizing problems can lead to difficulties with attention, concentration, memory, work completion, and problem solving, which can negatively influence academic performance. Furthermore, depression and anxiety can lead to avoidance of social situations, few interpersonal relationships, and weak quality of interpersonal relationships, which can affect social development (Huberty, 2014).

Assessment of Internalizing Problems

There are multiple methods for assessing internalizing problems, such as anxiety and depression, in youth. The most common methods are clinical interviews, direct observations, and rating scales (Gray et al., 2009; Klein, Dougherty, & Olino, 2005; Nardi, 2007; Silverman & Ollendick, 2005). These methods each come with strengths and weaknesses with respect to obtaining useful, accurate information about internalizing symptoms in youth or agreement across different sources.

Clinical Interviews. Clinical interviews are a commonly used method of assessment in child and adolescent psychology (Silverman & Ollendick, 2005). These interviews can be unstructured, semi-structured, or fully structured, with the format of the interview often depending on clinician preference (Klein, Dougherty, & Olino, 2005). Unstructured interviews have the most variability from clinician to clinician and may lead to a failure to inquire about key aspects of psychopathology; semi-structured interviews provide clinicians with a set of questions to ask but also allow for more flexibility with regard to asking additional follow-up questions for clarification or further inquiry; and structured interviews provide the least amount of flexibility as the clinician is limited to asking a pre-determined set of questions and recording responses. Semi-structured interviews may be used by more highly trained mental health professionals while more structured interviews are common when clinicians are less experienced and earlier on in their training (Klein, Dougherty, & Olino, 2005). Although clinical interviews are commonly used in the assessment of anxiety and depression, and can provide strong information (Gray et al., 2009), there can also be considerable variability in information obtained due to interview strategies and structure of the interview (Silverman & Ollendick, 2005). Additionally, even clinical interviews may result in poor agreement among multiple raters (i.e., the interviewers) because informants (i.e., the interviewees) may access accurate, but different information based on varying contextual facets of a child's behavior (De Los Reyes & Kazdin, 2005).

Direct Observation. Direct observation is another method for assessing internalizing problems in youth, and one of the main tools used for the assessment of behavioral, social, and emotional problems in children and adolescents within the school setting (Whitcomb & Merrell, 2014). Direct observation often requires more time and effort than other methods of assessment;

however, it is important for objectively assessing behavior (Hagopian & Jennett, 2014; Whitcomb & Merrell, 2014). With direct observation, the evaluator does not have to rely on retroactive reports from various informants; rather, the evaluator can observe the behavior directly and, when using a structured observation measure with behavioral definitions, obtain a more objective measurement of the behavior. For example, per the observable behaviors listed in the DSM-5 (APA, 2013), assessors may observe anxious individuals having trouble separating from their parents; failing to speak in social situations; being avoidant of specific objects, situations, or social interactions; experiencing panic attacks; and/or having problematic/irritable behavior. Additionally, individuals with depressive symptoms may appear tearful or sad, have diminished interest or pleasure in most activities, present with psychomotor agitation or retardation, or show signs of a diminished ability to think or concentrate (APA, 2013).

Although direct observation has the potential of being a more objective way to measure behavior, the accuracy, validity, and reliability of observational data may not be adequately established (Whitcomb & Merrell, 2013). These issues can be due to definitions of the behavior being too broad or too narrow, observer drift (i.e., observers gradually drift from original definitions of behavior), differences in observer training and reliability, observer reactivity, situational specificity (i.e., the child may behave differently across environments), and/or lack of comparison data (Whitcomb & Merrell, 2013). Further, internalizing disorders can be difficult to detect through external observation (Merrell, 2008), as many of the more prominent internal symptoms may be accessible only to the affected individual.

Rating Scales. Rating scales are among the main components of an assessment battery (Whitcomb & Merrell, 2013) for assessing behavioral, social, and emotional concerns; specifically, they are commonly used for screening and as part of the diagnostic process for

internalizing problems, such as anxiety and depression (Hagopian & Jennett, 2014). They are typically self-administered questionnaires completed by multiple informants that focus on current or recent symptoms and behavior, and they are a standardized and objectively scored method of measuring perceptions of behavior (Whitcomb & Merrell, 2013). Rating scales are often viewed as an efficient or cost effective way to assess symptoms and, therefore, are often used as screening instruments when there are behavioral concerns. When scores are elevated, a more comprehensive evaluation is typically completed; however, if informants do not accurately report symptoms, false negatives or false positives can occur (Klein et al., 2005). Perceived advantages of rating scales for assessment include: (a) cost effectiveness, (b) brevity (i.e., can learn a lot about the problem in a short amount of time), (c) can provide a summary of rater perceptions over time of the child or adolescent's behavior in natural environments (e.g., home or school), and (d) allow one to obtain data from individuals (e.g., parents, teachers, caregivers, etc.) who are most familiar with the child or adolescent and her/his behavior (Whitcomb & Merrell, 2013). There are, however, some disadvantages with rating scales as assessment tools, which are detailed in what follows below.

Parents and Teachers as Third-Party Informants. Parents and teachers are common informants when there are concerns about a child or adolescent's behavior or emotional difficulties. However, agreement between these two types of raters can vary. More overt behaviors (i.e., externalizing behavior) lead to stronger levels of agreement, whereas behaviors that are less obvious and involve more internal experiences (i.e., internalizing problems) lead to lower levels of agreement (Kanne et al., 2009; McDonald et al., 2016; Ung et al., 2017).

Discrepancies in parent and teacher ratings of child and adolescent behavior may be, in part, due to general measurement issues. Whitcomb & Merrell (2013) outlined two important

types of measurement issues related to rating scales: bias of response and error variance. Bias of response can occur based on the way the informants complete the rating scale and includes the halo effect, tendency toward leniency or severity, and/or central tendency of effects. The halo effect and negative bias refer to an informant rating a student more positively or negatively on an item because of separate positive or negative qualities the student has that are not related to the item being rated; leniency or severity is related to an informant rating students in an overly generous or overly critical manner; and a central tendency effect is the inclination for informants to endorse ratings in the middle of the scale (e.g., "sometimes" or "often") as opposed to selecting more extreme rating scale points (e.g., "always" or "never").

Error variance is another major issue concerning use of multiple informants completing rating scales as a method of assessment. The four main types of variance that may contribute error are source variance, setting variance, temporal variance, and instrument variance (Whitcomb & Merrell, 2013). First, source variance is characterized by the different types of response bias or response sets, described above, that can occur with different informants and the way they provide ratings. Next, setting variance is related to environmental differences in behavior. That is, a student may behave a certain way in school, but not at home—or the other way around. Such differences in behavior across contexts may be conceived of as a type of systematic error, but can also be viewed as a reflection of very real differences in behavior across settings. Additionally, temporal variance refers to the possibility that behavior ratings may be inconsistent over time either because the child/student's behavior changes or because the informant's approach to rating changes. Finally, instrument variance is related to different rating scales measuring slightly different constructs (e.g., "anxious/depressed" vs. "total anxiety"), using different normative samples to make score comparisons, and utilizing different descriptive

category cut-offs so that a behavior may fall in the "clinically significant" range on one measure and the "at risk" or "high-average" range on another (Whitcomb & Merrell, 2013).

It is important to obtain information from a variety of informants, sources, and methods in order to fully understand a child's functioning or symptom presentation across different naturalistic settings and to assist in determining appropriate interventions (Kanne et al., 2009; Taylor et al., 2018). However, previous research findings indicate that discrepancies among informants often exist, even when rating a child's behavior using the same instrument, which could be due to differences in behavior across time and setting (Kanne et al., 2009). If level of agreement among raters is weak, it can be challenging to integrate information in order to make clinical judgements (McDonald et al., 2016). These interpretation difficulties can lead to inaccurate diagnoses and decreased treatment efficacy (Ung et al., 2017).

Use of Self-Report Rating Scales. Internalizing problems, such as anxiety and depression, are more covert and may be less apparent or observable to third-party raters than externalizing behaviors; therefore, self-report assessment data are considered an essential addition to the assessment of internalizing problems (Merrell et al., 2002). However, because of their internal or less accessible nature, the measurement of emotions and moods involve unique challenges. Specifically, to obtain valid assessments, participants must first detect and integrate information regarding their internal experiences and then accurately convey those experiences (Humrichouse et al., 2007). Integrating information about internal experiences involves the coordination of multiple complex processes such as monitoring one's own psychological states, recognizing internal and external cues of emotions, and using emotion-related language (Durbin, 2010). This may be difficult for young children, especially children with ASD, to do given less

developed theory of mind and executive functions in these populations (Baron-Cohen, Leslie, and Frith, 1985; Hill, 2004; Kiep & Spek, 2016; Spek, Scholte, & Van Berckelaer-Onnes, 2009).

Self-Report Within ASD Samples. Assessing internalizing problems by relying on self-report could be potentially problematic in youth with ASD. For example, as characteristic of their diagnosis, youth with ASD tend to have difficulties with self-awareness and emotional understanding (Baron-Cohen, 2002; Lopata et al., 2010; Mazefsky et al., 2011); thus, they may find it difficult to accurately report their feelings and other internal states. Similarly, youth with ASD may have limited communication skills, which can impair their ability to self-report, convey, or otherwise communicate thoughts and emotional states (Hagopian & Jennett, 2014). Further, these difficulties that children and adolescents with ASD often have (e.g., demonstrating insight into emotions, accurately describing their own feelings, understanding what an item on a rating scale is asking, etc.) could mask or interfere with the detection of possible comorbid symptoms and diagnoses of internalizing problems such as anxiety and depression (Hammond & Hoffman, 2014).

Method and Measurement Issues in Cross-Informant Agreement

This section deals with method and measurement issues concerning inter-rater/cross-informant agreement at both the individual study level and the meta-analytic level (where studies are pooled). Issues at both levels are important within this literature.

Correlation Between Raters. In the meta-analytic context, the correlation coefficients from across studies would typically be transformed into Fisher's *z* prime scale values, averaged, and then the mean Fisher's *z* prime value would be back-transformed into the correlation coefficient metric. In general, this approach is considered to result in less systematic bias in the mean correlation outcome than would typically occur when raw correlation coefficients

themselves are averaged (Corey, Dunlap, & Burke, 1998). Note that Hedges and Olkin (1985) recommended using the Fisher's *z* prime method in the meta-analytic context, while Hunter and Schmidt (1990) argued for the use of the average of untransformed correlation coefficients, which is still a standard frequently used in psychometric meta-analyses (see Borenstein, Hedges, Higgins, & Rothstein, 2009; Hunter & Schmidt, 2004). Ultimately, Corey, Dunlap, and Burke (1998) showed that while averaged *r* and back-transformed *z* prime both showed evidence of bias, in general the back-transformed Fisher's *z* prime value results in less bias "under conditions common in meta-analysis" (p. 255). Thus, outside of psychometric meta-analysis, the Fisher's *z* prime method is preferred in order to minimize bias across the range of sample sizes.

Within the ASD literature, correlational agreement (reflecting covariation or association of individual scores or ratings across raters) has typically been reported in terms of the Pearson product moment correlation coefficient (e.g., Chow, 2008; Farrugia & Hudson, 2006; Hurtig et al., 2009; Kanne, Abbacchi, & Constantino, 2009; Lane, Paynter, & Sharman, 2013; Lopata et al., 2010; McDonald et al., 2016; Pisula et al., 2017; White, Schry, & Maddox, 2011).

However, a minority of studies have used a different index (e.g., intra-class correlation coefficient [ICC; see Blakeley-Smith et al., 2012; Jepsen, Gray, & Taffe, 2012; Magiati et al., 2014; Ooi et al., 2016; Ozsivadjian, Hibberd, & Hollocks, 2014; Kaat & Lecavalier, 2015; Ung et al., 2017] or Spearman's rho (see Hurtig et al., 2009). In general, the ICC represents a more precise measure of agreement because it takes into account both relative position and differences in means across the rater distributions (Liu, Tang, Chen, Lu, Feng, & Tu, 2016). In contrast, the Pearson product moment correlation coefficient is a measure of association reflecting similarity of position relative to each rater mean separately, while Spearman's rho reflects similarity of rank or ordinal position across the two raters (Rebekić et al., 2015)—and neither are sensitive to

mean differences between raters. It is not clear to what degree and under what circumstances it is legitimate to pool estimates of Pearson's *r*, Spearman's rho, and the ICC together as part of the same overall estimate in a meta-analysis. However, this is frequently done in practice (e.g., see Achenbach, Krukowski, Dumenci, & Ivanova, 2005; Achenbach, McConaughy, & Howell, 1987; Huang, 2017; Stratis & Lecavalier, 2015). In such instances, the variance of the ICC appears to have been calculated as if it were a Pearson correlation (i.e., via Fisher's *z* prime). In general, it appears that probability bias resulting from transforming the ICC (i.e., using Fisher's *z* prime) is smaller when the ICC reflects agreement between two groups of raters—as opposed to more than two groups (see McGraw & Wong, 1996). Two rater groups or rater-pairs is the most common inter-rater agreement context in this inter-rater literature (e.g., rater-pairs such as youth-parent, youth-teacher, parent-teacher, etc.).

Mean Differences Between Raters. The average level of rater scores or average amount of disagreement is often characterized via the mean difference between raters or rater types (see meta-analysis by Huang, 2017). In the meta-analytic context, this would require a common standardization metric to render mean differences comparable across studies, especially across studies using different rating scales or measurement instruments. A potential methodological issue to overcome is that some studies of average differences between raters may use an independent samples model of inference while others may use a dependent samples inferential model that takes the correlation between raters into account. If sufficient studies are available that take the dependency directly into account, then one could potentially use *dz* (i.e., mean difference between raters divided by the standard deviation of rater differences; see Cohen, 1988 p. 48) as the standard effect size metric. However, doing so would not allow inclusion of studies that used an independent samples model and did not provide the correlation between raters. This

problem suggests that the strategy for capturing and pooling the most mean differences using a common effect size metric would be to use Cohen's *d* or effect size *d* based on independent sample values (difference between means divided by the pooled standard deviation within; see Borenstein, Hedges, Higgins, & Rothstein, 2009). The dependent samples *dz* version can be readily converted into *d* if the correlation between raters is known. Because of the slight upward bias in *d*, especially in smaller samples, *d* values are typically adjusted with the resulting effect size value referred to as Hedges *g* (see correction formula in Hedges, 1981). Thus, within the ASD inter-rater literature, effect size *g* is most likely to allow for the pooling of the most studies and has the added benefit of correcting for the upward bias in *d*. Effect size *g* was the effect size estimate of choice in mean-level agreement or disagreement in Huang's (2017) meta-analysis of CBCL cross-informant agreement.

Discrepancy Scores. There appear to be three major types of discrepancy scores calculated for rater-pairs in the literature. These are (a) the difference between the raw or unstandardized ratings of the two informants, (b) the difference between the standardized ratings of the two informants, and (c) the residual difference between the two informant ratings (De Los Reyes & Kazdin, 2004; 2005). The raw or unstandardized difference method is calculated by simply subtracting one rater's unmodified score from the other. The difference between standardized ratings method initially requires raw scores for each rater type to be converted into *z* scores and then *z* scores for one rater are subtracted from the other. Finally, the residual difference method involves predicting one rater's score based on the other and then calculating the difference between the predicted rating and the actual rating.

It may be most important to distinguish between these types of discrepancy scores when attempting to correlate a discrepancy score with one or more other variables. As De Los Reyes

and Kazdin (2004) demonstrated in their sample, the particular discrepancy score calculation method can greatly influence the value of correlations with other variables. Thus, it may only make sense to compare correlations that involved use of the same type of discrepancy score. Otherwise, differences in the correlation coefficients could be at least partially due to an artifact (i.e., difference in method of discrepancy score calculation). In general, when a discrepancy score calculation method is required in the correlational context, De Los Reyes and Kazdin (2004) recommended use of the standardized difference score method, because it yielded more balanced correlational values with the two rater scores used to calculate it, yielded more consistent correlation estimates between discrepancy scores and informant demographics, and was "statistically discernable from the informants' ratings from which it was created" (p. 334). Whether this issue would matter at the descriptive level in the inter-rater or cross-informant context would depend on whether or not a researcher is interested in the correlation between a discrepancy score (i.e., some type of difference between two scores) and one or more external variables (e.g., age, years of education, etc.). Otherwise, descriptive indexes within the inter-rater or cross-informant agreement context typically involve correlations between the scores produced by different raters (i.e., the correlation between the two scores that make up a difference score) or the mean difference across two different raters or rater types--to which this issue does not apply.

Broad Theoretical Considerations for Cross-Informant Agreement

The Attribution Bias Context (ABC) Model. De Los Reyes and Kazdin (2005) proposed a general theoretical model of cross-informant agreement, the ABC Model, to guide research on informant discrepancies and to be applied in explaining aspects of cross-informant agreement in clinical settings (see Figure 1). The authors indicate that the ABC Model allows

for the conceptualization of why informant discrepancies exist, taking into account contextual factors and differences among different rater-pairs. Specifically, this model includes research on the actor-observer phenomenon (i.e., observers of another person's behavior tend to attribute their behavior to dispositional qualities and fail to consider contextual factors), perspective taking on memory recall (i.e., the influence that a person's perspective of the event influences their memory recall to support their views), and source monitoring (i.e., how individuals make attributions for how they acquire memories of events). This theory suggests that informants can enter the clinical assessment process with varying motivations and, as such, may have different perspectives regarding the severity of the behavior (De Los Reyes & Kazdin, 2005). This theoretical model provides a potential explanatory framework for understanding why informant discrepancies may occur when rating child behavior.

Theoretical Considerations Regarding Self-Report in Youth with ASD

Given the processes involved in accurate self-report of internalizing states (i.e., recognize, understand, integrate, and describe internal experiences) and the associated features/characteristics of ASD, it is reasonable to consider if children and adolescents with ASD experience challenges with the reporting of internalizing problems such as anxiety and depression. At least three theoretical models/constructs, related to ASD, would predict difficulties with the self-report of internal states in those with ASD. These theoretical positions involve the presence of alexithymia, delayed development in theory of mind, and deficits in executive functioning.

Alexithymia. Alexithymia is characterized by difficulty in identifying and describing personal experiences of emotions (Heaton et. al., 2012). Individuals with alexithymia may be aware that they are experiencing an emotion; however, they may not be able to identify, describe,

or articulate that emotion (Bird & Cook, 2013). Bird and Cook indicated that accurate self-reporting requires a degree of emotional awareness. (Because of this, it may be important to use tools that do not rely on or assume accurate self-report or introspective awareness when measuring alexithymia itself). These authors also reported prevalence rates of alexithymia to be approximately 10% in the general population and between 40% and 65% in ASD samples. The characteristics of alexithymia include impairment in identifying and describing one's emotions, and the presence of alexithymia suggests a likely weakened ability to self-report concerning internalizing issues (Bird & Cook, 2013). Because alexithymia frequently co-occurs with ASD, this may partially explain lower self-report of anxiety and depression within ASD relative to third party ratings.

Theory of Mind. Spek, Scholte, and Van Berckelaer-Onnes (2009) described theory of mind as the ability to attribute mental states to self and others. Baron-Cohen, Leslie, and Frith (1985) reported that children with autism fail to employ a theory of mind. Now, it is better understood that most children with ASD experience impairment or delayed development in inferring their own mental states (Spek, Scholte, and Van Berckelaer-Onnes, 2009). Happé (2003) also suggested that individuals with ASD may experience deficits in self-reflection. More specifically, some individuals with an ASD may have delays in the development of the cognitive processes that represent their thoughts and feelings as thoughts and feelings. The delayed development of a theory of mind would likely make it difficult for individuals with ASD to understand and report their emotions or other internal experiences.

Executive Function. Executive function is the broad term that encompasses narrower functions such as planning, working memory, impulse control, inhibition, and self-monitoring; these functions are all linked to the prefrontal cortex and involved in working memory, cognitive

flexibility, and inhibitory control (Hill, 2004). There is evidence that executive function deficits are present in ASD; specifically, that individuals with ASD tend to show impaired mental control that reduces one's ability to utilize problem solving for future planning (Kiep & Spek, 2016). Additionally, executive functions may support the development of theory of mind (Pellicano, 2012), implying that deficits in executive functioning can also lead to impairment in the development of theory of mind. Further, weakened self-monitoring (i.e., difficulty monitoring one's own thoughts and actions) related to deficits in executive functioning may be present in individuals with ASD (Hill, 2004). This difficulty with self-monitoring could make it challenging to recognize, and later report on, internalizing problems.

Prior Meta-Analyses

Prior meta-analyses have investigated broad cross-informant agreement of emotional/behavioral issues, assessed prevalence rates of anxiety and depression in individuals with ASD, examined cross-informant correlation coefficients or mean-level differences (not both in ASD), and included samples of youth with ASD and/or other comorbid disabilities (e.g., included studies with ASD and studies of ID without ASD). The meta-analyses reviewed here provide some useful information about the issues of interest in the present meta-analysis (i.e., cross-informant agreement and internalizing symptoms in youth with ASD), but critical examination also highlights clear gaps in the literature that speak to the need for presently proposed study.

The Achenbach et al. (1987) meta-analysis of 119 studies summarized cross-informant correlations across a variety of informant pairs (e.g., involving combinations of parents, teachers, mental health workers, observers, peers, and children/adolescents). Effect size estimates were derived from all available prior studies involving cross-informant ratings of children/adolescents.

Thus, estimates were calculated from a broad array of child sample types (e.g., typically developing children, children with various emotional and behavioral issues, children receiving inpatient services, children receiving outpatient services, etc.). Achenbach et al. (1987) found that correlations among all types of rater-pairs were significant, although correlations were higher between similar raters (i.e., parent vs. parent) than different raters (i.e., range of *r* between similar rater types = 0.54 to 0.74; range of *r* between different rater types = 0.20 to 0.44). Further, this meta-analysis revealed that correlations were generally higher for "undercontrolled" (i.e., externalizing) problems ($r = 0.41$) than "overcontrolled" (i.e., internalizing) problems ($r = 0.32$; see Table 1). Overall, the Achenbach et al. (1987) study provided useful information about multi-informant agreement patterns across a diverse array of youth samples. That is, it provided estimates of cross-informant agreement in a broad, general sense. However, it did not provide comparisons of cross-informant agreement correlations across different clinical conditions—other than broadly conceived internalizing and externalizing issues. As previously indicated, there are theoretical reasons for suspecting that these agreement estimates may differ for those with ASD (Baron-Cohen, 2002; Bird & Cook, 2013; Hagopian & Jennett, 2014; Kiep & Spek, 2016; Lopata et al., 2010; Mazefsky et al., 2011; Spek, Scholte, & Van Berckelaer-Onnes, 2009;). In addition, the Achenbach et al. (1987) study was completed several decades ago—leading to questions about how well its conclusions generalize to current rating instruments and current practice.

Achenbach et al. (2005) completed a meta-analysis of 108 studies examining cross-informant (e.g., spouses, family members, peers, clinicians, etc.) correlations of psychopathology in adults. Results of this study yielded mean cross-informant correlational levels of $r = 0.44$ for externalizing problems and $r = 0.43$ for internalizing problems when using the same instrument

to assess behavior (see Table 1). However, the authors of this meta-analysis state that they only included studies that utilized adult samples of participants who did not have conditions that would restrict their functioning (e.g., autism or IQs below 50). Therefore, it does not clearly contribute to understanding cross-informant agreement, regarding emotional/behavioral issues, among youth with ASD and, particularly, between self-report ratings by such youth and third-party reporters.

Stratis and Lecavalier (2015) conducted a meta-analysis of cross-informant agreement regarding emotional and behavioral problems, as well as social skills, in youth with either ASD or an ID. The authors state that the average correlational value across informants (i.e., parents, teachers, and children) and domains (i.e., emotional and behavioral problems and social skills) was in the moderate range (i.e., *r* values ranging from 0.25 to 0.71) with higher overall agreement for externalizing problems ($r = 0.42$) compared to internalizing problems ($r = 0.35$), and a stronger overall correlation between similar raters (i.e., raters of the same type; e.g., parent-parent or teacher-teacher; $r = 0.64$) than across different types of raters (e.g., youth-parent, parent-teacher, or teacher-youth; $r = 0.33$). More specifically, for externalizing problems, youth-parent association was $r = 0.44$; youth-teacher association was $r = 0.34$; parent-teacher $r = 0.38$. For internalizing problems, youth-parent association was $r = 0.42$; youth-teacher association was $r = 0.25$; and parent-teacher $r = 0.25$ (see Table 1).

Although this study provided insight concerning multi-informant agreement of internalizing problems among those with ASD or ID, its inclusion of studies that focused on youth with an intellectual disability in the absence of ASD made the results of this meta-analysis less meaningful for better understanding cross-informant agreement of internalizing symptoms exclusively among youth with ASD. It appears likely that the authors combined studies of these

two partially overlapping (partially comorbid) conditions to increase the number of relevant studies for the meta-analysis—contributing to overall statistical power and improving the likelihood of meeting minimum requirements for some moderator analyses. However, this sample of studies does not allow one to isolate the relationships within ASD and given the diversity of conditions that could involve ID, some might theoretically be expected to yield stronger agreement than in ASD. That is, the net effect would be averaging over potentially important differences.

In addition, Stratis and Lecavalier (2015) pooled effect size estimates across broad internalizing measures and measures of narrower internalizing constructs (e.g., anxiety, depression, etc.) into one overall internalizing estimate. Again, this strategy potentially allows for the inclusion of more studies with benefits of increased statistical power and improved options for moderator analyses. However, this approach conflates narrower constructs with a broader construct—with consequences of potentially missing differences in effect size between narrower constructs, as well as narrower constructs likely failing to more fully represent the broader construct. It is possible that narrower aspects of internalizing yield the same degree of cross-informant agreement as seen in broader internalizing, but this should not be something assumed without prior evidence to support it.

Huang (2017) investigated correlations and mean differences on the CBCL across a broad range of youth studies (n = 169) involving typically developing and clinical samples (ages 6-18 years). Informants included in this meta-analysis were parents, teachers, and the youths themselves (i.e., self-report). The correlations among informants were found to be small to moderate. For example, for the internalizing scale on the CBCL, the correlation between teachers and youth was r = 0.19; correlation between parents and youth was r = 0.33;

correlations between parents and teachers was $r = 0.18$. For the externalizing scale on the CBCL, the correlation was $r = 0.32$ between teachers and youth, $r = 0.40$ between parents and youth, and $r = 0.35$ between parents and teachers. The findings of this meta-analysis provided evidence to support stronger agreement between parent and youth raters than between other rater-pairs on internalizing and externalizing problems with higher agreement for externalizing problems compared to internalizing problems (see Table 1). Additionally, for mean differences, the effect size between parents and youth was $g = -0.21$ with youth reporting more internalizing problems than parents; the effect size for teachers and youth was $g = -0.76$ with youth reporting more symptoms than teachers; the effect size between parents and teachers was $g = 0.52$ with parents reporting more internalizing problems than teachers. Again, this meta-analysis provides useful information about multi-informant agreement on internalizing problems and incorporates clinical samples of youth. However, the sample is not specifically categorized and therefore does not evaluate multi-informant agreement data from only samples of youth with ASD.

Van Steensel and Heeman (2017) examined whether anxiety levels were elevated in children with ASD. The results of this meta-analysis of 83 studies revealed that when compared to other clinically referred youth (fixed model $d = 0.23$; random model $d = 0.12$) and typically developing youth (fixed model $d = 0.78$; random model $d = 0.97$), children with ASD had higher levels of anxiety. Additionally, the authors found the difference to be positively associated with IQ. This meta-analysis provided evidence that children with ASD are at generally higher risk for developing anxiety disorders (see Table 1)—using a mean of the parent and child effect sizes to represent an average effect size for that study. However, it did not address cross-informant agreement regarding anxiety disorders or ratings of anxiety in this population of youth—either from an inter-rater correlation perspective or a mean difference between raters perspective

Hudson et al. (2018) explored the prevalence of depressive disorders in children, adolescents, and adults with ASD. This meta-analysis of 66 studies found current and lifetime prevalence of depressive disorders in children 18 years and under to be 10.6% and 7.7%, respectively. In adults 18 years and over, current prevalence was found to be 19.4% with lifetime prevalence being 40.2%. Analyses indicated studies that (a) used standardized interviews to assess for the depressive disorders, (b) asked participants to report on their own depressive symptoms, and (c) included participants with higher intelligence yielded the highest rates of depressive disorders (see Table 1). Similar to the Van Steensel and Heeman (2017) meta-analysis of anxiety in ASD, this study provided useful information about the prevalence of depression in individuals with ASD. However, also like Van Steensel and Heeman, the Hudson et al. (2018) study did not provide data relating to cross-informant agreement.

Finally, Hollocks et al. (2018) also conducted a meta-analysis of 35 studies (27 measuring anxiety, 29 measuring depression, and 21 measuring both) examining prevalence rates of anxiety and depression in adults with ASD. This study found that the lifetime prevalence estimate for an anxiety disorder in adults with ASD was 42%, while the lifetime prevalence of a depressive disorder in adults with ASD was 37%. Current prevalence rates were 27% for an anxiety disorder and 23% for depression in this population. Although Hollocks et al. (2018)'s findings indicated that individuals with ASD are at a higher risk of developing comorbid mental health conditions (see Table 1), this meta-analysis did not address any multi-informant agreement issues. It also focused on studies of adults with ASD; consequently, it's findings are less applicable to the child and adolescent context.

Research on the extent of agreement between self-report and third-party reports (e.g., parent report, teacher report, etc.) of anxiety, depressive, and broad internalizing symptoms

within the population of youth with ASD is mixed and incomplete. Some studies have found that youth with ASD and third-party reporters have acceptable agreement when reporting levels of internalizing problems (Farrugia & Hudson, 2006; Ozsivadjian, Hibberd, & Hollocks, 2014). On the other hand, many other studies have found poor cross-informant agreement between youth with ASD and parents or teachers, wherein youth with ASD reported lower levels of anxiety and depression than parents or teachers (Bitsika & Sharpley, 2015; Barnhill et. al., 2000; Kaat & Lecavalier, 2015; Lopata et. al., 2010; White, Schry, & Maddox, 2012). Although some published meta-analyses explore related topics, there are currently no meta-analyses that have investigated multi-informant agreement in anxiety, depression, and broad internalizing problems specifically in youth with ASD. Understanding how reports of such symptoms may differ across raters is critical to making sense of whether, and to what degree, reliance on different rater types across different studies may lead to different results.

It is crucial to identify, as accurately as possible, symptoms of anxiety, depression, and broad internalizing problems in youth with ASD. This is important for establishing clinical need, reducing the likelihood of missed cases, and receiving appropriate interventions in clinical and/or school settings as early as possible. Though inter-rater agreement does not guarantee accuracy, consistency and extent of agreement and disagreement across rater types are critical pieces of information for providing interpretive context in both applied and research settings. Ultimately, understanding the extent to which ratings, and other information, provided from different sources relate to broadly accepted external criteria for these constructs is the long-term goal. In the absence of such gold standard, broadly accepted, external indicators, agreement across sources (treating each as predictor and criterion relative to the other) is a critically important preliminary step—with important interpretive implications. Although prior meta-analyses have provided

useful information about informant agreement regarding child emotional and behavioral problems (e.g., broad clinical and typical samples, expressed in terms of correlation, in youth [Achenbach et al., 1987] and adults [Achenbach et al., 2005]; broad clinical and typical samples involving the CBCL, expressed as correlation and mean differences [Huang, 2017]; broad externalizing, internalizing, and social skills pooled across studies of ASD and/or ID, and expressed in terms of correlation [Stratis & Lecavalier, 2015]; etc.), and provided related information concerning internalizing issues (e.g., prevalence of anxiety and depression in adults with ASD [Hollocks et al., 2018]; prevalence of depressive disorders in children, adolescents, and adults with ASD [Hudson et al., 2018]; and group-level differences in anxiety between ASD and other broad clinical or typically developing groups [Van Steensil & Heeman, 2017]), there are important elements that have not yet been explored (e.g., agreement in cross-informant ratings specifically within ASD—in terms of correlation and mean differences for anxiety, depression, and broad internalizing constructs). In order to further investigate and expand on this important topic, the current meta-analysis will focus on samples of youth with ASD specifically, examine narrower internalizing constructs (i.e., anxiety and depression) in addition to broader internalizing, and investigate cross-informant agreement conceived of as both correlations and mean differences.

Inter-rater agreement is important for the identification of co-occurring internalizing symptoms in ASD. If correlational values are weak and mean levels are significantly different among raters, it provides an unclear summary of symptomology that can lead to false negatives or false positives of internalizing symptoms. For example, relying on self-report that yields lower levels of symptoms could lead to more false negatives whereas relying on parent report that indicates high levels of symptomology could lead to more false positives. Multi-method and

multi-source assessments are recommended as standard practice (Taylor et al., 2018); however, poor inter-rater agreement could make the work of interpreting and synthesizing information, regarding internalizing problems, across multiple sources and methods more difficult in youth with ASD. If poor agreement is present, it is possible that depending on what construct, measurement tool, and rater/reporter is used when assessing for internalizing symptoms, studies could yield very different estimates of symptomology. Differences between mean levels across rater types (i.e., youth, parent, teacher) may ultimately be important in determining which rater tends to produce more useful information, which may further depend on the context. The more we understand about the potential differences that are likely to occur, based on different sources and measurement methods for operationalizing the same construct, the better prepared we will be to select or favor assessment approaches most relevant to the context and interpret findings with a realistic level of confidence.

Potential Moderators

Age of Youth. There is evidence to suggest that youth with ASD can vary in their ability to self-report based on age and cognitive abilities (Vasa et al., 2016). Specifically, within a typically developing sample of young children (ages 3 to 6 years), there was greater convergence between children's reports of emotions and objective coding of expression when the children were older, had higher verbal intelligence, and greater emotion-regulation abilities (Durbin, 2010). Similarly, in a clinical sample of youth with depression, anxiety, and/or conduct problems, younger children's self-reports and parent reports did not significantly correlate on internalizing measures; however, older children's self-reports and parent reports of internalizing problems did significantly correlate (Ebesutani et al., 2011). Thus, older age lead to more agreement between self-report and parent report ratings on internalizing measures. Within a

sample of youth with either ASD or intellectual disability, as age increased, so did the level of agreement among different informants (e.g., youth-parent, youth-teacher, parent-teacher, etc.) on internalizing problems; therefore, the authors interpreted that youth are more accurate self-reporters as they get older (Stratis & Lecavalier, 2015). Achenbach et al. (1987) found that, across articles examining cross-informant relationships involving a variety of different sample types, multi-informant correlations were significantly higher for adolescents aged 12 to 19 years (r = 0.51) than for children aged 6 to 11 years (r = 0.41). In general, research indicates that youth age can moderate the relationship among rater-pairs; thus, age of youth will be explored as a moderator in this meta-analysis.

Cognitive Ability of Youth. As mentioned above, research supports greater agreement between youth self-report and objective measures of behavior expression when children had higher verbal intelligence scores and more advanced cognitive abilities (Durbin, 2010; Vasa et al., 2016). Specifically, studies have found better agreement between parents and youth with ASD to be associated with higher youth IQ (verbal IQ and full scale IQ) and more advanced cognitive skills (Blakeley-Smith et al., 2011; Kaat & Lecavalier, 2015; Ooi et al., 2016). This could be partially about children having a broader vocabulary of emotion-related words leading to greater convergence between their subjective experience of emotions, verbal labeling of emotions, and more objective measures (Durbin, 2010). Given this prior evidence from other contexts, this meta-analysis will explore youth cognitive ability as a moderator variable.

Method of Self-Report Administration. Another potential moderator may be the method of self-report administration. That is, where the self-report rating scales were completed (e.g., in the clinic/school, research, or home setting), and how the self-report rating scales were administered (e.g., youth completes unassisted, clinician/researcher reads items to all youth in

sample, clinician/researcher reads items to youth as needed, parent reads items to all youth in sample, parent reads items to youth as needed/at their discretion, etc.) may be related to agreement.

A number of the studies that found higher levels of agreement between parents and child raters were those in which the self-report rating scale was administered at home (Bitsika et al., 2015; Farrugia & Hudson, 2006; Jepsen, Gray, & Taffe, 2012; Magiati et al., 2014; Ozsivadjian et al., 2013). In most cases, studies only reported that the questionnaires were mailed to homes, completed by parents, youth, (and at times, teachers), and sent back to the researchers (e.g., Farrugia & Hudson, 2006; Jepsen et al., 2012; Magiati et al., 2014; etc.). However, other studies provided more details about specific instructions given to participating families. For example,, Bitsika et al. (2016) reported that researchers told parents to be in the vicinity of their child when the child completed the rating scale—in case the child became anxious, but not close enough to see the responses. Additionally, Ozsivadjian et al. (2013) indicated that when questionnaires were sent home in the mail, parents were asked to assist their child with reading the items if needed, but not guide their responses; however, when families did not complete the questionnaires before their first appointment, the children completed them at the clinic.

In general, studies where youth completed self-report questionnaires in the clinic with items read to them by researchers appeared to find poorer youth-parent agreement (e.g., Chow, 2008; Lopata et al., 2010; Taylor et al., 2018; etc.). When youth complete self-report rating scales at home, it is possible that parents may inadvertently influence their child's ratings or youth may ask parents for assistance determining which rating to give a particular item. This would undermine the assumption of independence between the raters. The lack of independence could increase inter-rater agreement under these circumstances. Although this has not been

investigated in previous studies, there is some evidence (as noted above) to support method of self-report administration as a moderator and, therefore, it will be explored in this meta-analysis.

Score Type. In the broader inter-rater agreement literature, studies vary in terms of whether agreement or discrepancy is assessed based on norm-referenced standard scores or raw scores. The potential implications of this distinction are not small, as effect size estimates may vary considerably depending on the score type used. For example, Huang (2017), in a meta-analysis of cross-informant agreement using the CBCL, found that standardized mean differences were generally larger for raw scores than for norm-references standard scores—and also found the direction of the effect size differed in the youth-parent (norm-referenced g = 0.30 vs. raw score g = -0.66) and youth-teacher (norm-referenced g = 0.42 vs. raw score g = -1.40) rater-pair contexts. Within the ASD inter-rater literature, most studies appear to report agreement or discrepancy information in relation to norm-referenced standard scores (e.g., Bitsika & Sharley, 2015; Blakeley-Smith et al., 2012; Chow, 2008; Davidsson et al., 2017; Farrugia & Hudson, 2006; Gillott et al., 2011; Lopata et al., 2010; Park et al., 2013), while a small number report only raw scores (e.g., Normann Andersen et al., 2015; Ozsivadjian et al., 2014; Pisula et al., 2017; Soloman et al., 2012; White et al., 2011). In such cases, it appears that raw scores were used because norm-referenced standard scores are/were not available for the particular measure involved or because the norms may not have been an appropriate or an applicable standard for the sample involved (e.g., Normann Andersen et al., 2015; White et al., 2012). In the case of comparing means across raters, an important issue for raw score use would be that the two rater forms contain comparable numbers of items, rated on the same scale, and with comparable item content. (Note that when norm-referenced standard scores are used, less than perfect item comparability across rater forms is less of an issue—especially when the

different rater forms were co-normed.) It would be difficult, if not impossible, to evaluate the possible meaning of mean differences between rater types without some standard for comparability across the rater forms involved.

Parent Socioeconomic Status (SES). A prior meta-analysis (Duhig et al., 2000) found that lower levels of interparental agreement were found for lower-SES parents than middle-SES parents. Although this finding is limited to a similar rater-pair (i.e., parent-parent), it provides some evidence to suggest that SES may affect agreement between other rater-pairs that include a parent (i.e., youth-parent; parent-teacher). However, De Los Reyes and Kazdin (2005), in their review, reported that the relationship between informant discrepancies and SES is inconsistent across studies and indicated that studies examining a variety of informant pairs have not found a relationship between SES and informant agreement when other child and parent characteristics were accounted for. Therefore, it is possible that findings related to parent SES moderating informant agreement are better accounted for by other variables. However, De Los Reyes and Kazdin's findings did leave open the possibility that SES may play a more unique moderating role under more narrow or special conditions.

Ethnicity/race. According to the review by De Los Reyes and Kazdin (2005), in the broader literature, it has been found that: (a) either agreement is lower or discrepancies tend to be greater across raters for African American youth compared to European American youth; (b) African American children tend to rate themselves as more anxious than their mothers rate them, while European American children tend to rate themselves as less anxious than their mother's ratings; and (c) a meta-analysis of cross-parent ratings of youths did not yield evidence of a relationship between ethnicity and cross-rater agreement (Duhig, Renk, Epstein, & Phares, 2000, as cited in De Los Reyes & Kazdin, 2005).

Gender. In the broader inter-rater literature, findings concerning gender differences have been inconsistent, with broad meta-analyses (e.g., Achenbach et al., 1987) typically finding no significant differences overall. According to a review by De Los Reyes and Kazdin (2005) "…the mixed findings from these investigators suggest that in aggregate, child gender may not be related to informant discrepancies, but in specific populations, child gender effects may be present" (pp. 485-486). Though gender differences in ASD have been reviewed and documented (e.g., Mandy et al., 2012; Van Wijngaarden-Cremers et al., 2014), no inter-rater studies examining potential relationships between inter-rater agreement/discrepancies and gender in ASD have been reported.

Social Desirability. Social desirability is discussed in the broader inter-rater literature primarily from the self-report perspective. The assumption being that children and adolescents would rate themselves more favorably or less pathological due to self-consciousness and/or perceived social desirability, and that this would partially account for discrepancies with other raters (e.g., Jensen, Xenakis, Davis, & Degroot, 1988). However, De Los Reyes and Kazdin (2005) found only mixed support for this hypothesis, and De Los Reyes et al. (2015) advised strongly against interpreting relatively lower self-ratings from youth compared to the ratings of others, in isolation, as necessarily reflecting the influence of social desirability. Given the differences in social perception within the ASD context, it is not clear that one would anticipate a systemic effect of social desirability on self-ratings to be as strong—even if more consistent findings were seen in the broader literature. However, this remains an open question. In addition, there is also the open question of to what extent social desirability might play a role in parent ratings of the youth's behavior.

Parental Depression. It has been suggested that parent depression may increase the likelihood of a negative bias on the part of the parent in perceiving a child's behavior or emotional issues. This is referred to as the depression-distortion hypothesis (Richters, 1992). According to reviews by De Los Reyes and Kazdin (2005) and De Los Reyes et al. (2015), support for the depression-distortion hypothesis is mixed and, when present, the influence in terms of incremental variance involved does not appear to be large enough to warrant typically discounting parent ratings in this situation.

Parental Stress. The review by De Los Reyes and Kazdin (2005) reported on several inter-rater studies suggesting a positive relationship between parent stress and cross-informant discrepancies. However, these authors also cautioned that it is not clear to what degree various aspects of parental psychopathology may account for overlapping aspects of the variance in cross-rater discrepancies (e.g., parent stress, parent depression, parent anxiety, etc.). Though this review examined studies outside of the ASD context, it is reasonable to think that this relationship between parent stress/psychopathology and cross-informant discrepancies would likely generalize to ASD. However, this remains an open question.

Moderator Availability. For purposes of conducting a meta-analysis of inter-rater agreement in youth with ASD, there are a number of considerations regarding these other possible moderator variables. First, based on a preliminary review of studies, variables such as youth age, youth cognitive ability, and method of self-report administration will likely be available in most cases and with reasonable variability across studies. Second, variables such as score type, gender, and ethnicity will also likely be reported across studies. However, it is likely that studies may not yield sufficient variability in these characteristics to assess them well as possible moderators of inter-rater agreement or cross-informant discrepancy. Gender differences

in ASD have been noted outside of the inter-rater context (e.g., Mandy et al., 2012; Van Wijngaarden-Cremers et al., 2014). However, most ASD studies tend to contain far more males than females with the gender ratio tending to vary by functional level (Volker & Lopata, 2008). Very few if any inter-rater studies are likely to involve an exclusively female sample or sufficiently large separate female group for comparison purposes at this time. Additionally, in the cases of social desirability, parent depression, and parent stress, it is unlikely that many studies in the ASD inter-rater literature report information on these variables as none of the studies in the preliminary review did so.

Need for Present Study

Findings in the literature related to inter-rater agreement of internalizing problems in ASD are variable and inconsistent. Currently, no prior meta-analysis has focused on the issue of inter-rater agreement in this population of children and there is no systematic examination of the similarities and differences (i.e., potential moderators) that may help to explain discrepant findings. This meta-analysis will add to the research on the topic of internalizing problems in youth with ASD by exploring inter-rater agreement and the potential moderator variables that may improve agreement, which will contribute important information on the utility, accuracy, and potential bias of ratings provided by multiple informants.

Research Questions and Hypotheses

Research Question 1: What is the relationship between different informants' ratings of anxiety, depression, or broader internalizing in youth with ASD? Rater-pairs consisting of youth-parent, youth-teacher, and parent-teacher will be examined and this research question will focus on the correlation (i.e., association) between the rater-pairs. However, it is anticipated that more articles will be available for anxiety (compared to depression or broad internalizing) and youth-

parent (compared to youth-teacher or parent-teacher). Additionally, prior research is scarce and inconsistent regarding the correlation between youth-teacher rater-pairs when reporting on anxiety, depression, or internalizing symptoms. Specifically, two studies (Hurtig et al., 2009; Jepsen et al., 2012) both measured these constructs in youth with ASD using the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach & Rescorla, 2001), but found different results. Across the two studies, correlation values between youth and teacher ratings ranged from .16 to .34 on the anxiety construct, from .09 to .66 on the depression construct, and from .06 to .56 on the internalizing problems construct. Therefore, given insufficient prior data, a specific hypothesis concerning these relationships within the youth-teacher rater-pair is not feasible at this time.

- Hypothesis 1a: The mean correlation effect size between youth self-report and parent report ratings of anxiety, depression, or broad internalizing in youth with ASD will be significant and yield a small to medium effect.

    Previous meta-analyses examining cross-informant agreement have found youth-parent correlations for broad internalizing problems to range from .33-.42 (Huang, 2017; Stratis & Lecavalier, 2015). However, Huang (2017) used a broad collection of CBCL samples (i.e., typical and broad clinical) and Stratis and Lecavalier (2015) examined studies involving youth with ASD and youth with ID (without ASD)—combined into one meta-analysis. There are theoretical reasons (i.e., possible alexithymia, delayed theory of mind, and executive function impairments) suggesting that self-report might be poorer or less accurate within ASD-specific samples—and potentially leading to lower agreement with other raters. A number of observational studies using ASD samples reported

correlations between youth-parent ratings of anxiety, depression, and/or broad internalizing problems. Specifically in this literature, youth-parent correlations for anxiety symptoms ranged from -.02 to .69 (Blakeley-smith et al., 2012; Chow, 2008; Lopata et al., 2010; Magiati et al., 2014; Ooi et al., 2016); youth-parent correlations involving depression symptoms ranged from .29 to .31 (Chow, 2008; Hurtig et al., 2009; Lopata et al., 2010); and youth-parent correlations concerning broad internalizing symptoms ranged from .25 to .56 (Hurtig et al., 2009; Jepsen et al., 2012; Kaat & Lecavalier, 2015). (It is possible that the diversity of correlation values may ultimately be accounted for by potential moderator variables [e.g., age of youth, cognitive ability of youth, method of self-report administration, score type, parent SES, social desirability, parental depression, or parental stress].)

- Hypothesis 1b: The mean correlation effect size between parent report and teacher report ratings of anxiety, depression, or broad internalizing in youth with ASD will be significant and yield a medium effect.

  An observational study conducted by Kanne et al. (2009) using an ASD sample found correlations between teacher and parent ratings of broad internalizing problems to be fairly low (i.e., affective scale $r = 0.08$; anxiety scale $r = 0.14$). However, McDonald et al. (2016) completed an observational study using a high-functioning ASD sample and found better (i.e., moderate to high) agreement between parent and teacher raters (i.e., internalizing problems scale $r = 0.28$; anxiety scale $r = 0.34$; depression scale $r = 0.30$). Jepsen et al. (2012) also used an ASD sample in their study and reported parent-teacher rater-pair correlations

to be low to moderate (i.e., anxious/depressed scale $r = 0.26$; withdrawn/ depressed scale $r = 0.35$; internalizing problems scale $r = 0.21$). Because this rater-pair does not include a youth rater with ASD, potential issues with self-report based on theory (i.e., possible alexithymia, delayed theory of mind, and executive function impairments) that may weaken informant agreement are not relevant.

Research Question 2: Is the correlation between different informants' ratings of anxiety, depression, or broad internalizing in youth with ASD moderated by the general cognitive ability of the youth? There is not enough evidence, based on prior studies or theory, to develop a specific hypothesis related to parent-teacher agreement in this context; however, hypotheses have been generated for youth-parent and youth-teacher agreement.

- Hypothesis 2a: The correlation between youth self-report ratings and parent ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the general cognitive ability of the youth, such that more advanced cognitive ability will be associated with better youth-parent agreement.

  In an observational study, Ooi et al. (2016) found better agreement between parents and youth with ASD in reporting anxiety symptoms was associated with higher youth verbal IQ. Similarly, Blakeley-Smith et al. (2011) determined that, in an ASD sample, higher youth-parent agreement was associated with higher verbal IQ and more advanced metacognitive skills. In another observational study with an ASD sample, youth-parent agreement was better for youth with higher IQs (Kaat & Lecavalier, 2015). In contrast, a meta-analysis by Stratis and Lecavalier (2015) found an inverse relationship between IQ and inter-rater

agreement across raters for broad internalizing. However, this meta-analysis combined studies of youth with ASD with studies of non-ASD youth with ID—and it is not clear how the resulting heterogeneity may have impacted the findings. This finding appears to be an anomaly and may be an artifact that arose from the different samples involved. Otherwise, in general, across broad samples of youth, the literature supports higher agreement between youth self-report and other measures of behavior occur when children have higher verbal or higher general cognitive ability (Durbin, 2010; Vasa et al., 2016).

- Hypothesis 2b: The correlation between youth self-report ratings and teacher ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the general cognitive ability of the youth, such that more advanced cognitive ability will be associated with better youth-teacher agreement.

  Based on the same evidence used to support hypothesis 2a, (Blakeley-Smith et al., 2011; Kaat & Lecavalier, 2015; Ooi et al., 2016), it is logical, by extension, that more advanced cognitive ability will be associated with better youth-teacher agreement. That is, higher youth cognitive ability is likely associated with better self-perception, which may improve agreement with third party raters who are assumed to know the youth well and be reasonable observers of the youth's behavior.

Research Question 3: Are the correlations between self-report and other informants' ratings of anxiety, depression, or broad internalizing in youth with ASD moderated by the age of the youth?

- Hypothesis 3a: The correlation between self-report and parent ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the age of the youth with older age leading to a stronger positive correlation.

  In a study involving a typically developing sample of children, young children's self-reports of anxiety, depression, and broad internalizing problems did not significantly correlate with parent reports (anxiety $r = 0.16$; depression $r = 0.008$; internalizing $r = 0.05$), whereas older children's self-reports did significantly correlate with parent reports (anxiety $r = 0.28$; depression $r = 0.16$; internalizing $r = 0.215$; Ebesutani et al., 2011). Similarly, Achenbach et al. (1987) found that, across a broad sample of youth studies, multi-informant correlations were stronger for older children ($r = 0.51$) than younger children ($r = 0.41$). Additionally, using a sample of youth with either ASD or ID without ASD, Stratis & Lecavalier (2015) found that level of agreement among raters on internalizing problems was positively associated with age. There are also some theoretical considerations (i.e., possibly better developed theory of mind and executive functioning skills with increasing age) that support more accurate self-reporting as all children age, which may lead to better agreement with adult raters.

- Hypothesis 3b: The correlation between self-report and teacher ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the age of the youth with older age being associated with a stronger positive correlation.

  As mentioned above, older children tend to have more developed theory of mind and executive functions potentially making it less difficult to understand, integrate, and convey their internal experiences (Hill, 2004; Kiep & Spek, 2016;

Spek, Scholte, & Van Berckalaer-Onnes, 2009). Additionally, as stated in hypothesis 3a, it has been found that as age of the child increases, so does the level of agreement among raters across both broad and ASD/ID samples of youth (Achenbach et al., 1987; Stratis & Lecavalier, 2015). Overall, older children are considered to be better at self-reporting their symptoms and, as a result, agreement with other reporters may increase.

- Hypothesis 3c: The correlation between parent ratings and teacher ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the age of the youth with older age leading to stronger positive correlation.

  Internalizing problems can present differently based on child age (Layne et al., 2009). For example, young children may either present with temper tantrums/misbehavior or very few noticeable symptoms (Frick et al., 1994), which could make it more difficult for third-party raters to detect those as internalizing symptoms. As such, parents and teachers may have different perspectives about the child's behavior, which can lead to lower agreement in ratings. Conversely, internalizing problems in older children may present more identifiably (e.g., withdrawn mood, irritability, worrying behaviors, etc.). This could then increase the agreement between parent and teacher raters when reporting on internalizing symptoms.

Research Question 4: Is the correlation between self-report and parent ratings of anxiety, depression, or broad internalizing in youth with ASD moderated by the method of self-report administration (e.g.,, assessment read to child by researcher/clinician in the clinic, assessment

read to the child by parent at home, assessment completed independently by the child in the clinic, etc.)?

- Hypothesis 4: The correlation between self-report and parent ratings of anxiety, depression, or broad internalizing will be moderated by the method of self-report administration with the conditions "assessment read to child at home" and "assessment completed at home" being associated with a stronger positive correlation.

    This higher correlation in situations where the parents may be involved in the administration is predicted based on the likelihood of one rater influencing the other by being made responsible for administering the rating instrument to the youth with ASD. In this case, there is the very real possibility that parents can influence the ratings of the youth—whether intentionally or not. Thus, the higher correlation in this situation may be an artifact resulting from a lack of clear independence between the ratings completed by youth and parent. Many studies that included ASD samples and involved self-reports being completed in the home setting (e.g., Bitsika et al., 2015; Farrugia & Hudson, 2006; Jepsen, Gray, & Taffe, 2012; Magiati et al., 2014) found high/acceptable levels of agreement between parent and self-report of internalizing problems. Conversely, studies where youth with ASD completed self-report measures in the clinic, with items read to them by a research clinician, often found poor agreement between youth and parent ratings (e.g., Chow, 2008; Lopata et al., 2010; Taylor et al., 2018).

Research Question 5: What are the mean differences between different informants' reports of anxiety, depression, or broad internalizing in youth with ASD? Rater-pairs consisting of youth-parent, youth-teacher, and parent-teacher will be examined. However, it is anticipated that more

articles will be available for anxiety (compared to depression or broad internalizing) and youth-parent (compared to youth-teacher or parent-teacher). This research question will focus on the corrected standardized mean difference (i.e., Hedges *g*) between rater-pairs.

- Hypothesis 5a: When rating the behavior of youth with ASD, mean parent-rated anxiety, depression, or broad internalizing scores will be significantly higher than mean youth self-report ratings.

    This hypothesis is supported by prior findings from observational studies (Barnhill et al., 2000; Bitsika & Sharpley, 2015; Kaat, 2014; Lopata et al., 2010; Taylor et al., 2018), which found that parents of children with ASD tend to report higher mean levels of anxiety, depression, or internalizing problems for the youth compared to the youth with ASD self-report. This hypothesis is also supported by theoretical positions suggesting that youth with ASD may be less accurate at self-reporting their internal symptoms (i.e., possible alexithymia, delayed theory of mind, and impairments in executive functions).

- Hypothesis 5b: When rating the behavior of youth with ASD, mean teacher-rated anxiety, depression, or broad internalizing scores will be significantly higher than mean youth self-report ratings.

    This hypothesis is supported by Barnhill et al. (2000). Their study found that, when compared to youth with ASD, teachers reported higher levels of anxiety (teacher $M = 60.10$; youth $M = 47.19$) and depression (teacher $M = 62.00$; youth $M = 50.56$) in the youth relative to youth self-report. As with hypothesis 5a, theoretical evidence for this hypothesis includes the possible presence of

alexithymia, less developed theory of mind, and executive function issues that youth with ASD may experience.

- Hypothesis 5c: When rating the behavior of youth with ASD, mean teacher-rated anxiety, depression, or broad internalizing scores will not differ substantially from mean parent ratings.

  McDonald et al. (2016; parent-rated anxiety *M [SD]* = 55.43 [12.59], teacher-rated anxiety *M [SD]* = 57.57 [12.49]; parent-rated depression *M [SD]* = 57.58 [12.05], teacher-rated depression *M [SD]* = 58.72 [11.15]; parent-rated internalizing *M [SD]* = 55.14 [11.59], teacher-rated internalizing *M [SD]* = 56.86 [11.19]) and Barnhill et al. (2000; parent-rated anxiety *M [SD]* = 59.60 [13.37], teacher-rated anxiety *M [SD]* = 60.10 [7.91]; parent-rated depression *M [SD]* = 69.40 [12.13], teacher-rated depression *M [SD]* = 62.00 [13.37]; parent-rated internalizing *M [SD]* = 65.10 [13.75], teacher-rated internalizing *M [SD]* = 60.90 [10.31]) both reported similar mean levels across parent and teacher raters on anxiety, depression, and internalizing constructs when rating symptoms in youth with ASD.

Research Question 6: Are the mean differences between informant's ratings of anxiety, depression, or broad internalizing in youth with ASD moderated by the general cognitive ability of the youth? Prior research does not provide enough support regarding the relationship between youth cognitive ability and mean level differences in parent and teacher ratings; thus, a reasonable hypothesis cannot be generated for this rater-pair. Nonetheless, specific hypotheses about youth-parent and youth-teacher mean level differences have been developed.

- Hypothesis 6a: Mean differences between youth self-report ratings and parent report ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by youth cognitive ability with more advanced cognitive ability being associated with a smaller mean difference.

    As discussed in hypothesis 2a, individual studies and the broader literature suggest that higher youth IQ is associated with greater youth-parent agreement in ASD and typically developing samples (Blakeley-Smith et al., 2011; Durbin, 2010; Kaat & Lecavalier, 2015; Ooi et al., 2016; Vasa et al., 2016). Though more of this research is directed at correlations than mean differences, the available evidence suggests it is reasonable to hypothesize that mean level differences will be smaller for youth with better cognitive abilities.

- Hypothesis 6b: Mean differences between youth self-report ratings and teacher report ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by youth cognitive ability with more advanced cognitive ability being associated with a smaller mean difference.

    Again, based on the prior research findings that support better youth-parent agreement in ASD and ASD/ID samples as youth IQ increases (Blakeley-Smith et al., 2011; Kaat & Lecavalier, 2015; Ooi et al., 2016). This requires some extrapolation from the youth-parent context, but it is reasonable, assuming that better cognitive ability is associated with or suggestive of improved self-perception, to expect a similar trend with youth-teacher agreement. Based on this assumption, mean level differences will be smaller for youth with more advanced cognitive abilities than youth with less developed cognitive abilities.

Research Question 7: Are the mean differences between informant's reports of anxiety, depression, or broad internalizing in youth with ASD moderated by the age of the child?

- Hypothesis 7a: Mean differences between youth self-report and parent report ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by youth age with older age being associated with a smaller mean difference.

  Per Huang (2017) who completed a meta-analysis regarding CBCL cross-informant agreement concerning behavior problems using a broad range of typical and clinical samples, the mean-level disagreement between parents and youths when reporting on internalizing problems was significant for younger children (ages 6-10 years: g = -1.07), but not significant for older children (ages 11-14 years: g = -0.12; ages 15-18: g = -0.15). Secondary evidence for this hypothesis includes findings (discussed above) that older child age in broad clinical and ASD/ID samples leads to a stronger correlation between parents and youth ratings (Achenbach et al., 1987; Stratis & Lecavalier, 2015)—although correlations and mean differences focus on different types of variation.

- Hypothesis 7b: Mean differences between youth self-report and teacher report ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by youth age with older age associated with smaller mean differences.

  As was discussed in Hypothesis 3b, older children tend to have better developed processes and functions that allow them to more easily understand and report on their internal experiences (Hill, 2004; Kiep & Spek, 2016; Spek, Scholte, & Van Berckalaer-Onnes, 2009). This may lead to increased accuracy in self-report for

youth with ASD and, thus, predict that mean differences between youth and teacher ratings will decrease as youth age increases.

- Hypothesis 7c: Mean differences between parent ratings and teacher ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by youth age with older age associated with a smaller mean difference.

  Using the same rationale that supported Hypothesis 3c, it can be expected that with regard to parent and teacher ratings, mean differences will be smaller in older samples of youth than they are in younger samples of children.

Research Question 8: Are the mean differences between informant's ratings of anxiety, depression, or broad internalizing in youth with ASD moderated by the method of self-report administration? This research question applies only to youth-parent rater-pairs.

- Hypothesis 8: Differences in mean scores between youth self-report and parent report of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the method of self-report administration with the conditions "assessment read to child at home" and "assessment completed at home" leading to a smaller mean difference.

  Based on previous findings from observational studies utilizing ASD samples (Bitsika et al., 2015; Farrugia & Hudson, 2006; Jepsen, Gray, & Taffe, 2012; Magiati et al., 2014), similar mean level ratings of anxiety, depression, and internalizing problems were found between parent report and youth self-report when the assessment was completed at home. However, in studies where items were read to youth by a research clinician in the clinic setting, more divergent mean differences between youth and parent ratings were found (e.g., Chow, 2008; Lopata et al., 2010; Taylor et al., 2018). When parents are involved in the

administration of the youth's rating instrument (i.e., reading items/assisting youth when needed) it is possible that they will inadvertently influence the youth's responses in a way that yields similar mean scores on the instrument.

Exploratory Analyses. Apart from the specific hypotheses above, correlations and mean differences between other rater-pairs will be assessed, as non-directional hypotheses, when sufficient data are available. In addition, other potential moderator variables (e.g., score type, socioeconomic status, parental depression, etc.) will be assessed providing a sufficient number of studies are available for analyses to be conducted. In cases where several studies are available but insufficient for a formal statistical analysis, the study results, similarities, and differences will be examined visually and conceptually for potential patterns, which though statistically inconclusive, may be suggestive for future studies to examine.

CHAPTER III

METHOD

Meta-analyses synthesize the available evidence for a research question using data from prior studies (Borenstein et al., 2009). The present study consists of a meta-analysis examining cross-informant (i.e., youth-parent, youth-teacher, and parent-teacher rater-pairs) agreement of anxiety, depression, and broad internalizing issues, and the potential moderators, in youth ASD samples through the investigation of both correlations and mean differences. The PRISMA meta-analysis guidelines (Moher et al., 2009) were used in the writing of this paper.

Literature Search

To ensure that a comprehensive set of articles were identified, a multi-step procedure was employed, which included computerized searches of relevant online databases. Potentially relevant search terms were searched in PsycINFO including PsycARTICLES, ERIC, PubMed/MEDLINE, and ProQuest Dissertations and Thesis. Additionally, content type included book/eBook, book chapter, dissertation/thesis, journal article, and manuscript. Publication date was not restricted in order to maximize the number of articles identified and prevent limiting findings to only more recent publications. The following search terms were used in this literature search:

1. autis* OR ASD* OR aspergers* OR HFASD
2. AND internaliz* OR anxiety* OR depress*
3. AND inform* OR rating* OR report* OR parent OR self OR teacher OR agree* OR disagree* OR converge* OR correspond*
4. AND assess* OR treat* OR measure*

Asterisks indicate that search terms with that root were used as a keyword. Following this search, an additional search included an examination of article and dissertation reference

lists. See Figure 2 for the PRISMA flow diagram outlining the number of articles identified, screened, coded, and included.

Abstract Screening Procedure

Article abstracts were screened to determine if they (a) were published in English, (b) utilized an ASD youth sample (i.e., age 18 years or under), (c) measured anxiety, depression, or internalizing problems, and (d) included youth self-report and parent report, youth self-report and teacher report, or parent report and teacher report (see Appendix A for abstract screening criteria checklist). Only abstracts that clearly did not meet criteria were screened out. Abstracts that meet criteria or potentially meet criteria were selected for full-text screening (k = 5,126; number of abstracts, articles, or studies will be denoted as k).

Full-text Screening Procedure

Full-text screening was completed for articles that met the above criteria following abstract screening (k = 402). These articles were screened using the following criteria (see Appendix B for criteria checklist):

1. The study includes one or more rating scales that assess depression, anxiety, or internalizing problems.
2. The study includes multiple informants, specifically at least one rater-pair. Target rater-pairs include youth-parent, youth-teacher, or parent-teacher.
3. Either (a) for each measure and each rater, the sample size, mean, and standard deviation were reported or (b) critical information about the correlation (association/agreement) between rater-pairs' scores is available (i.e., sample size for the rater-pair and correlation coefficient).

Data Coding

If criteria was met during the full-text screening stage, articles were then thoroughly coded using a coding sheet (see Appendix C). Generally, each study (k = 107) was coded for the following information:

1. *General article information:* article name, study authors, publication year, state/country in which the study was conducted, journal name, type of publication (e.g., journal article, book or book chapter, dissertation, thesis, unpublished report, etc.), study design, type of informant/rater-pairs, diagnoses in ASD group (e.g., ASD, HFASD, Asperger's Syndrome, etc.), and how the diagnoses were made.
2. *Group specific demographic information (ASD and comparison group):* age range, age mean, age standard deviation, ethnicity, gender, and (if reported) participant IQ (range, mean, and standard deviation), adaptive functioning, language functioning, social desirability, parent depression, and parent stress (conveyed in terms of either the categorical or continuous metric available in the study).
3. *Group specific measurement information:* measure used for parent report, self-report, and/or teacher report; and for each measure, the construct, score type (e.g., total, subscale, etc.), sample size, mean score, standard deviation of score, and descriptive category (if reported). For self-report only, articles were coded for method of self-report completion (e.g., assessment read to child by researcher or clinician in clinic, assessment read to child at home by parent, assessment completed without assistance in clinic, assessment completed at home reportedly without assistance, or other [allowing for parents to help children complete the assessment, but not requiring it, completed at home, but no specification of boundaries on parent assistance, etc.]).

4. *Group-specific and rater-pair-specific correlational information:* measure and construct, sample size of the rater-pair, correlation coefficient, mean difference, and standard deviation of paired difference (if available) or standard deviations for different rater types involved.

Protecting Reliability

Each independent coder ran the literature search using the above search terms and screened the abstracts and full-text articles using the criteria checklists (see Appendix A and Appendix B). Then, the coders each coded the agreed upon articles (see Appendix C for coding sheet). Among the coders, an overall agreement of 95% was achieved in the abstract screening phase, 96% during the full-text screening stage, and 100% in the coding phase.

Rating Scales Used to Measure Internalizing Constructs

The included studies used a variety of rating scales to measure the internalizing constructs of interest (i.e., anxiety, depression, and broad internalizing problems) across studies. The most commonly utilized rating scales were: The Multidimensional Anxiety Scale for Children (MASC; March, 1997); Children's Depression Inventory (CDI; Kovacs, 2003); Behavior Assessment System for Children, Second Edition—Parent Rating Scale (BASC-2-PRS; Reynolds & Kamphaus, 2004), Self-Report of Personality (BASC-2-SRP; Reynolds & Kamphaus, 2004), and Teacher Rating Scale (BASC-2-TRS; Reynolds & Kamphaus, 2004); Achenbach System of Empirically Based Assessment (ASEBA)—Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), Youth Self Report (YSR; Achenbach & Rescorla, 2001), and Teacher's Report Form (TRF; Achenbach & Rescorla, 2001); Spence Children's Anxiety Scale (SCAS; Spence, 1998); Screen for Child Anxiety Related Disorders (SCARED; Birmaher et al., 1995); and the Revised Children's Anxiety and Depression Scale (RCADS; Chorpita et al.,

2000). A more detailed description of each of these anxiety, depression, or broader internalizing measures is given below.

Each study can be represented by only one effect size estimate per construct within the correlational analysis and within the mean difference analysis. In cases where one study includes multiple measures of the same construct (e.g., depression measured using the BASC-2 and CDI), certain rating instruments were prioritized over others for use in calculating effect size estimates. Priority was given to instruments that result in the best match across rater types (e.g., co-normed BASC-2-PRS and BASC-2-SRP) or instruments that are more commonly used among the other studies included in the meta-analysis (to minimize variation across studies due to instrument differences). Commonly included measures are as followed:

Multidimensional Anxiety Scale for Children (MASC). The MASC (March, 1997; and second edition [MASC-2], March, 2012) was designed to screen for anxiety symptoms in children ranging in age from eight to 19 years old and includes a parent and self-report form. Each item is rated using a four-point Likert scale (0 - Never true about my child [me], 1 - rarely true about my child [me], 2 - sometimes true about my child [me], and 3 - often true about my child [me]), with higher ratings indicating greater anxiety severity. For composites and subscales, norm-referenced *T*-scores that are 60 or above suggest elevated symptoms of anxiety. The MASC provides a Total Anxiety score, which has acceptable internal consistency and test-retest reliability (March & MHS Staff, 1997; March, Parker, Sullivan, Stallings, & Conners, 1997). Only the MASC Total Anxiety Score will be used in this meta-analysis.

Children's Depression Inventory (CDI). The CDI (Kovacs, 2003; and second edition [CDI-2], Kovacs & MHS Staff, 2010) is a parent and self-report instrument that assesses depression in children ages seven to 17 years old. Items are rated on a scale from 0 to 2 (0 =

absence of symptom; 1 = mild symptom; 2 = definite symptom) based on what was experienced in the last two weeks. Higher scores on the CDI indicate greater depression severity with scores above 35 (range = 0-54) suggesting symptoms of depression. The CDI Total Score, which has acceptable reliability and validity (Kovacs, 2003), will be included in this meta-analysis.

Behavior Assessment System for Children, Second Edition (BASC-2). The BASC-2 (Reynolds & Kamphaus, 2004; note also the original BASC [Reynolds & Kamphaus, 1994; and the third edition [BASC-3], Reynolds & Kamphaus, 2015) is a behavior assessment system used to evaluate symptoms in individuals ages two to 25 years old. On the BASC-2-Parent Rating Scale (PRS) and BASC-2-Teacher Rating Scale (TRS), items are rated using a four-choice frequency response scale (0 = Never, 1 = Sometimes, 2 = Often, and 3 = Almost Always). On the BASC-2-Self-Report of Personality (SRP), items are rated using a *True* or *False* response or a four-choice response scale (0 = Never, 1 = Sometimes, 2 = Often, and 3 = Almost Always). Higher scores on the clinical scales are indicative of more severe behaviors. Specifically, norm-referenced *T*-scores 70 or above on the clinical scales are considered Clinically Significant. The BASC-2 has appropriate reliability and validity (Reynolds & Kamphaus, 2004). The anxiety, depression, and internalizing problems scales for the BASC-2-PRS, BASC-2-SRP, and BASC-2-TRS will be included in this meta-analysis.

Achenbach System of Empirically Based Assessment (ASEBA)—Child Behavior Checklist (CBCL), Youth Self Report (YSR), and Teacher's Report Form (TRF). The ASEBA (Achenbach & Rescorla, 2001) is a comprehensive behavior assessment system that evaluates behavioral functioning in individuals ages one and a half to 90+ years and has acceptable reliability and validity (Achenbach & Rescorla, 2001). On the CBCL, YSR, and TRF, items are rated using a Likert scale (0 = Not True, 1 = Somewhat or Sometimes True, 2 =

Very True or Often True). Higher scores represent more emotional severity (*T*-scores 70 and above indicate Clinical levels of behavior symptoms). Anxious/Depressed (anxiety construct), Withdrawn/Depressed (depression construct), and Internalizing Problems scales from the CBCL, YSR, and TRF will be included in this meta-analysis.

Spence Children's Anxiety Scale (SCAS) and Spence Children's Anxiety Scale-Parent (SCAS-P). The SCAS/P (Spence, 1998) is a rating scale aimed at assessing for specific anxiety disorder symptoms in children. Specifically, the six subscales of the SCAS/P are Social Phobia, Panic Disorder, Agoraphobia, Generalized Anxiety Disorder, Obsessive Compulsive Disorder, Separation Anxiety Disorder, and Specific Phobias; the SCAS/P also includes a Total Score. The child self-report scale consists of 44-items and the parent report scale contains 38-items; for both versions, items are rated using a 0 (Never) to 3 (Always) point scale. Higher scores on the SCAS/P are indicative of greater severity of anxiety symptoms with *T*-scores above 60 suggesting elevated levels of anxiety. The SCAS/P has good internal reliability, test-retest reliability, convergent/divergent validity, discriminant validity, and construct validity (Spence, 1998). The Total Score on the SCAS/P will be utilized in this meta-analysis.

Screen for Child Anxiety Related Disorders (SCARED). The SCARED (Birmaher et al., 1995) is a parent and self-report measure used to screen for signs of anxiety disorders in children. Each item is rated using a 3-point scale ranging from 0 (Not True or Hardly Ever True) to 2 (Very True or Often True) with higher scores suggesting more severe anxiety. Specifically, a raw score of 25 or above suggests the presence of an anxiety disorder. The SCARED total score will be included in this meta-analysis and has been found to demonstrate moderate youth-parent agreement and good internal consistency and discriminant validity (Birmaher et al., 1999).

Revised Children's Anxiety and Depression Scale (RCADS) and Revised Children's Anxiety and Depression Scale-Parent Version (RCADS-P). The RCADS/P (Chorpita et al., 2000) is a questionnaire used to assess symptoms of anxiety and depression in children. Items included in this measure are rated using a 4-point Likert-scale ranging from 0 (Never) to 3 (Always). Higher scores indicate more severe symptomology with *T*-scores 70 or higher reaching the Clinical range. Along with subscales corresponding to DSM-IV-TR diagnoses (i.e., Separation Anxiety Disorder, Social Phobia, Generalized Anxiety Disorder, Panic Disorder, Obsessive Compulsive Disorder, and Major Depressive Disorder), the RCADS/P yields a Total Anxiety Scale score and a Total Internalizing Scale score. Both of these total scores have been found to have good reliability and validity (Chorpita, Moffitt, & Gray, 2005) and will be included in this meta-analysis.

Data Analyses

Correlational Effect Size. When studies report a correlation between two continuous variables, it is appropriate for the correlation coefficient to serve as the effect size index (Borenstein et al., 2009). Therefore, for the studies included in this meta-analysis that reported level of agreement (i.e., more precisely association, assessed via correlation) between multiple rater-pairs on symptoms of anxiety, depression, or broad internalizing, the correlation coefficient was used as the effect size index. Pearson correlations (k = 47), intra-class correlations (k = 18), and Spearman's *rho* (k = 11) were reported in the studies included in this meta-analysis. Following Stratis and Lecavalier's (2015) approach, intra-class correlations and Spearman's *rho* correlations were treated as Pearson correlations in order to maximize the number of studies available for analyses. Of note, type of correlation coefficient was analyzed as a moderator variable to investigate whether the type of correlation coefficient made a difference in the results.

The correlation coefficients (Pearson correlations, Spearman's *rho*, or intra-class correlations) were converted to the Fisher's *z* scale for averaging and confidence interval calculations (see formulas below, ln = natural logarithm; *r* = correlation coefficient; *n* = sample size; V_z = variance of *z*. Following this process, the summary values were back transformed into the correlation metric (Borenstein et al., 2009; see below).

Correlation transformed to Fisher's *z* scale:

$$z = 0.05\ x \ln(\frac{1+r}{1-r})$$

Variance of *z*:

$$V_z = \frac{1}{n-3}$$

Standard error:

$$SE_z = \sqrt{V_z}$$

Conversion from Fisher's *z* scale back to a correlation:

$$r = \frac{e^{2z}-1}{e^{2z}+1}$$

Mean Difference Effect Size. When possible, the standardized mean difference between the raters in each pair were calculated to assess the standardized level of disagreement from an average level/mean difference perspective. Thus, the standardized mean difference was also used as an effect size index in a second set of analyses. The standardized mean difference was chosen because the studies that will be included in this meta-analysis used different instruments to measure levels of anxiety, depression, and/or broad internalizing (unstandardized mean differences can only be used when all of the studies in the meta-analysis use the same scale for measuring the outcome variable[s]). Hedges' *g* was used for effect size estimates. To calculate a standardized mean difference for each outcome variable, the mean difference was divided by the

pooled standard deviation within, creating an index that is comparable across studies (Borenstein et al., 2009). Then, *d* values were converted to Hedges' *g*, using the correction factor *J*, to remove the slight bias of Cohen's *d* (see below for calculation and conversion equations; $\bar{X}_1$ and $\bar{X}_2$ = sample means in the two groups; n_1 and n_2 = sample sizes in the two groups; s_1 and s_2 = standard deviations in the two groups).

Computing the standardized mean difference:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\mathrm{SD}_{within}}$$

$$S_{within} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Variance of *d*:

$$V_d = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}$$

Standard Error of *d*:

$$SE_d = \sqrt{V}_d$$

Conversion from *d* to Hedges' *g*:

$$J = 1 - \frac{3}{4df - 1}$$

Effect sizes for each rater-pair (i.e., youth-parent; youth-teacher; parent-teacher) and each outcome construct (i.e., anxiety, depression, and/ or internalizing problems) were treated independently. However, when studies included multiple effect sizes in the same behavior category for the same rater-pair using different rating scales, the most comparable measure across rater-pairs was used (e.g., BASC-2-SRP vs. BASC-2 PRS as opposed to BASC-2-SRP vs. RCADS-P). Consistent with Stratis and Lecavalier (2015), if multiple comparable measures are

used, the effect size from the more commonly used rating scale (compared to the other studies included in the meta-analysis) was included in order to increase consistency across comparisons.

Correlational Analysis. An overall weighted average correlation estimate and associated 95% confidence interval was determined for each rater-pair under each relevant research question and/or hypothesis. In calculating each overall mean correlation, the effect size from each contributing study was weighted according to its inverse variance. Mean correlation estimates were tested for significance using the *Z* test statistic. Potential heterogeneity among the effect size estimates contributing to each average correlation estimate was tested for significance using the *Q* statistic—or more specifically excess variation quantified as $Q - df$ was tested for statistical significance, while the ratio of excess variation to total variation was assessed via reported I^2 values. (Other variation indices, such as *T* [standard deviation of effect size estimates] and T^2 [variance of effect size estimates], were examined for additional interpretive context.) Care was taken to consider if the significance test (based on $Q - df$) may be insensitive (due to too few or imprecise studies) or overly sensitive (due to many studies or the presence of many precise estimates) depending on the number of studies and characteristics of the studies involved. The associated I^2 statistic helps determine if sufficient excess variation is present that would support the need for moderator analyses—in an attempt to account for the excess variation in effect estimates across studies.

Mean Difference Analysis. The same general set of procedures was followed for the bias-corrected standardized mean difference effect size (Hedges *g*; Hedges, 1981) estimates—reflecting differences between the means of different pairs of raters relative to their pooled standard deviation. An overall weighted average effect size *g* estimate and associated 95% confidence interval was determined for each rater-pair under each relevant research question

and/or hypothesis. When calculating each overall mean *g* estimate, the effect size from each contributing study was weighted according to its inverse variance and mean effect size *g* estimates were tested for significance using the *Z* test statistic. Potential heterogeneity among the effect size estimates contributing to each average *g* estimate were tested for significance using the *Q* statistic (excess variation quantified as $Q - df$ was tested for statistical significance, while the ratio of excess variation to total variation was assessed via reported I^2 values). Other variation indices, such as *T* and T^2 were also examined. As in the case with the correlation effect size estimates, factors that could impact the sensitivity of the heterogeneity significance test (based on $Q - df$) were considered. The associated I^2 statistic was also examined to determine if sufficient excess variation was present. If a study did not report correlational values along with means and standard deviations, the average correlation calculated for each category in the correlational analyses was used as the correlation coefficient for the study.

Analysis of Publication Bias. It is important for the articles included in a meta-analysis to be distributed properly in order to accurately represent the topic of interest. A potential distribution issue is the file-drawer problem, which refers to the tendency for articles with statistically significant results to be preferentially published (Pautasso, 2010). This reflects the placing of less value on studies with findings that are not statistically significant; however, these results (i.e., the lack of statistical significance) can also provide very useful information for determining the real or true range of effect size variation across studies. To determine if publication bias has potentially played a significant role in the sample of studies available for and included in this meta-analysis, funnel plots were constructed. If publication bias is not present, distributions will appear symmetrical about the mean effect size (appearing in the shape

of a funnel); if publication bias is likely present, there will be less consistent symmetry with gaps in the plot (Borenstein et al., 2009).

If bias is present, it is important to explore whether the bias had any impact on the observed effect or if it was entirely responsible for the observed effect. To do this, Rosenthal's Fail-safe *N* method can be employed to compute the number of missing studies that would need to be included before the *p*-value became nonsignificant (Borenstein et al., 2009). If this method reveals that a large number of studies would be needed in order to nullify the observed effect, there is less reason for concern; however, if only a small number of studies would be needed, there should be more concern. The Fail-safe *N* statistic is the number of missing studies that would need to be included in the current study to overturn the rejection of the null hypothesis (Howell & Shields, 2008).

Computational Model. The computation model that is selected should be chosen based on the expectation of whether the studies share a common or "true" underlying effect size. Specifically, a fixed-effects model would be appropriate if all the studies included in the meta-analysis are functionally identical and if the goal is to calculate a common effect size for the population. Conversely, a random-effects model should be selected when there is no reason to assume that the studies included are identical and have the same true effect size. Specifically, when using a random-effects model, it is assumed that the true effect size varies across studies (Borenstein et al., 2009). The studies selected for this meta-analysis operated independently and had varying background attributes (e.g., location, researchers, characteristics of ASD sample, etc.), preventing them from being functionally equivalent; thus, a random-effects model was used to estimate the mean effect size of each study. This was calculated using the equations below (V_{Y_i}= the within-study variance for study *i*; *k* = number of studies; Y_i = observed effect).

Determining the weight assigned to each study:

$$W_i = \frac{1}{V_{Y_i}}$$

Computing the weighted mean:

$$M = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i}$$

Variance of the summary effect:

$$V_M = \frac{1}{\sum_{i=1}^{k} W_i}$$

Standard error of the summary effect:

$$SE_M = \sqrt{V_M}$$

Null hypothesis that mean effect is zero:

$$Z = \frac{M}{SE_M}$$

Further, a mixed-effects model can be used if one or more moderators are included in the model that may account for some heterogeneity and would investigate the extent to which the moderators included impact the true effect (Viechtbauer, 2010). Moderators were included in the analyses when possible to attempt to account for any excess variation in effect size estimates across studies. Thus, a mixed-effects model was used to determine the impact that moderator variables may have in accounting for variation using the following equation (θ_i = the unknown true effect; u_i = the average true effect of the study; x_i = value of the moderator for the study).

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{p'} x_{ip'} + u_i$$

General Analytic Strategy. All calculations and analyses were performed using the Comprehensive Meta-Analysis Version 3 (Borenstein, Hedges, Higgins, & Rothstein, 2014) software program. Of major interest is the degree of association and mean differences within

each type of rater-pair (i.e., youth-parent; youth-teacher; parent-teacher) for each internalizing behavior category (i.e., anxiety, depression, and/or broad internalizing). Potential moderators such as youth cognitive ability, youth age, method of self-report administration, and correlation type were also considered. Based on the type of information available across studies, cognitive ability (*M* FSIQ score) and age (*M* age in years) were analyzed as continuous variables. These potential moderators were analyzed using meta-regression with effect size as the outcome variable and the moderator as the predictor. Because method of self-report administration and correlation type are both a categorical variables, these potential moderators were dummy-coded to indicate the method of administration (i.e., 1 = assessment read to child in clinic; 2 = assessment read to child at home; 3 = assessment completed in clinic; 4 = assessment completed at home) or type of correlation coefficient (i.e., 1 = Pearson's *r*; 2 = ICC; 3 = Spearman's rho) and then analyzed using meta-regression.

CHAPTER IV

RESULTS

Overview of Studies Included

Seventy-five studies met criteria and were included in the current meta-analysis. In terms of effect size estimates, 38 studies were included in the correlation analyses and 60 studies were included in the mean difference analyses. Of the 75 total studies included, some studies were used for both correlational and mean differences analyses, accounting for some overlaps in numbers of studies included within each construct/rater-pair. The paragraphs that follow include information regarding studies included in the current study's analyses and are separated by effect size type (i.e., correlational vs. standardized mean difference).

Correlational Studies. Within the correlational analyses, 29 total studies (*n* per study range: 19—150 participant pairs) were included in the parent vs. self rater-pair category (anxiety: *k* = 25, depression: *k* = 16, broad internalizing: *k* = 5); 12 studies (*n* per study range: 22—177) were included in the parent vs. teacher rater-pair category (anxiety: *k* = 7, depression: *k* = 7, broad internalizing: *k* = 9); and three studies (*n* per study range: 22—36) were included in the teacher vs. self rater-pair category (anxiety: *k* = 2, depression: *k* = 3, broad internalizing: *k* = 2).

For moderator analyses within the anxiety construct for the parent vs. self rater-pair, 13 studies were included for cognitive ability of youth (Mean Full Scale IQ per study [*M* FSIQ] range: 87.78—110.14, *Mdn* = 101.66), 24 studies were included for age of youth (*M* age per study range: 9.75—15.06 years, *Mdn* = 12.65), and 13 studies were included for method of self-report administration. Within the depression construct for the parent vs. self rater-pair, 12 studies were included for cognitive ability of youth (*M* FSIQ range: 90.70—110.14, *Mdn* = 103.97), 15 studies were included for age of youth (*M* age per study range: 9.30—15.06 years,

Mdn = 13; 14 studies with one outlier removed for follow-up analyses), and seven studies were included for method of self-report administration. For the broad internalizing construct within the parent vs. self rater-pair, four studies were included for cognitive ability of youth (*M* FSIQ range: 90.70—103.94, *Mdn* = 93.02) and five studies were included for age of youth (*M* age per study range: 11.20—15.06 years, *Mdn* = 13). There were not enough studies available to perform moderator analyses for method of self-report administration within this construct (*k* =4).

Within the anxiety construct for the parent vs. teacher rater-pair, six studies were included for age of youth moderator analysis (*M* age range: 4.20—15.06 years, *Mdn* = 8.12), but there were not enough studies available for moderator analyses regarding FSIQ (*k* = 3). For the depression construct within the parent vs. teacher rater-pair, four studies were included for cognitive ability of youth (*M* FSIQ range: 81.30—105.41, *Mdn* = 97.13) and seven studies were included for age of youth (*M* age range: 4.20—15.06 years, *Mdn* = 8.75). Within the broad internalizing construct for the parent vs. teacher rater-pair, four studies were included for cognitive ability of youth (*M* FSIQ range: 81.30—103.12, *Mdn* = 92.50) and nine studies were included in age of youth (*M* age range: 4.20—15.06 years, *Mdn* = 7.47).

Moderator variables within the teacher vs. self rater-pair category were not able to be conducted due to insufficient studies available. The number of studies available ranged from two to three across the three constructs.

Mean Difference Studies. Within the mean difference analyses, 49 studies (*n* range: 6—170 participant pairs) were included in the parent vs. self rater-pair category (anxiety: *k* = 46; depression: *k* = 18; broad internalizing: *k* = 4), 18 studies (*n* per study range: 6—403) were included in the parent vs. teacher rater-pair category (anxiety: *k* = 11; depression: *k* = 9; broad

internalizing: *k* = 14), and six studies (*n* per study range: 6—36) were included in the teacher vs. self rater-pair category (anxiety: *k* = 6; depression: *k* = 4; broad internalizing: *k* = 2).

For moderator analyses within the anxiety construct for the parent vs. self rater-pair, 25 studies were included for cognitive ability of youth (*M* FSIQ range: 88.07—122.25, *Mdn* = 101.66), 42 studies were included for age of youth (*M* age range: 9.18—16.81 years, *Mdn* = 12.13), and 21 studies were included for method of self-report administration. Within the depression construct for the parent vs. self rater-pair, 12 studies were included for cognitive ability of youth (*M* FSIQ range: 90.70—122.25, *Mdn* = 102.91), 16 studies were included for age of youth (*M* age range: 9.33—15.06 years, *Mdn* = 12.05), and 12 studies were included for method of self-report administration. For the broad internalizing construct within the parent vs. self rater-pair, there were not enough studies available to complete moderator analyses for any of the three moderator variables.

Within the anxiety construct for the parent vs. teacher rater-pair, seven studies were included in the cognitive ability of youth (*M* FSIQ range: 72.70—122.25, *Mdn* = 97.94) moderator analysis and nine studies were included for age of youth (*M* age range: 4.30—15.06 years, *Mdn* = 10.70). For the depression construct within the parent vs. teacher rater-pair, six studies were included in the cognitive ability of youth (*M* FSIQ range: 72.70—122.25, *Mdn* = 94.54) and seven studies were included for age of youth (*M* age range: 4.30—15.06 years, *Mdn* = 8.74) moderator analyses. Within the broad internalizing construct for the parent vs. teacher rater-pair, seven studies were included in the cognitive ability of youth (*M* FSIQ range: 83.85—122.25, *Mdn* = 93.86) and 12 studies were included in age of youth (*M* age range: 4.30—15.06 years, *Mdn* = 9).

For the anxiety construct within the teacher vs. self rater-pair, four studies were included in the cognitive ability of youth (*M* FSIQ range: 91.13—122.25, *Mdn* = 101.69) and five studies were included for age of youth (*M* age range: 10.70—15.06 years, *Mdn* = 13.20) moderator analyses. Moderator variables for the depression and broad internalizing constructs within the teacher vs. self rater-pair category were not able to be conducted due to insufficient studies available (*k* range: 1—3).

General Approach to Research Questions

Results are reported under each general research question according to the order of the hypotheses, with relevant results that fall outside of specific hypotheses reported immediately following the related hypotheses (e.g., results concerning correlations for the teacher vs. self-ratings, which did not involve specific hypotheses, will immediately follow the results of correlations for parent vs. self-ratings and parent vs. teacher ratings—for which specific hypotheses were made). The results of any remaining exploratory analyses, that do not relate directly to particular research questions, are reported after the results of all of the research questions. For purposes of interpreting effect size estimates in what follows, Cohen's (1988) standards for correlation coefficients and standardized mean differences will be used, with each standard value being interpreted as the minimum value required to meet that standard. For correlation coefficients, $r = 0.10$ is the minimum standard for a small effect, $r = 0.30$ is the minimum standard for a medium effect, and $r = 0.50$ is the minimum standard for a large effect. For a standardize mean difference (e.g., Cohen's *d* or Hedges' *g*), 0.20 is the minimum for small, 0.50 is the minimum for medium, and 0.80 is the minimum for a large effect size (Cohen, 1988).

It is important to note up front that for hypotheses regarding correlations, the coefficients reported across studies were typically Pearson correlations—and all correlations were treated as

such for purposes of the analyses. However, several studies reported Spearman's *rho* or an ICC value for their correlation estimates. Combining across different types of coefficients is not ideal, but was done for the following reasons: (a) the number of studies involved was often not large and correlation values reported do reflect the available obtained values in the literature; (b) Stratis and Lecavalier (2015) included non-Pearson correlations in their effect size estimates, as this represented only a minority of studies, maximized the number of studies that could be included, and did not appear to substantively impact results; and (c) results of moderator analyses, to be reported later (using correlation coefficient type as a moderator), were non-significant—and associated effect size estimates for differences between the correlation types did not appear to be substantive.

The approach taken for reporting the moderator analyses requires some up-front description and explanation. Results for all moderator analyses include the following three component *Q* tests and associated descriptive indices. First, the *Q* test for the total true between study variance is reported. A significant result is generally interpreted as the presence of substantive true variability in effect size estimates beyond what would be expected based on sampling variation alone. Tau squared (T^2), Tau (T), and I^2 values are reported with this result. T^2 reflects the estimated true variance in effect sizes, T is the estimated standard deviation of the true effect sizes, and I^2 is an estimated percentage of the overall observed variance that is true variation in effect size estimates. This *Q* test is typically conducted prior to moderator analyses to determine if sufficient true variation in effect size estimates is present for potential moderators to account for. (Though non-significance may suggest lack of heterogeneity among effect size estimates, and limited need for potential moderators, this should be interpreted with caution due to the possible influence of sample size on the power of the test.) Second, the *Q* test for the

moderator meta-regression model is reported, along with the associated R^2_{Analog} value and the meta-regression equation where X = the proposed moderator. Third, a final Q test is reported that reflects whether significant variance in true effect size estimates still remains after accounting for the moderator. The T^2, T, and I^2 values are reported along with this Q test for residual variance. (Note that the R^2_{Analog} value reported with the second Q test is calculated by subtracting the T^2 value of the third Q test from the T^2 value of the first Q test, and then dividing the result by the T^2 value from the first Q test. At times, the reported R^2_{Analog} value may be slightly discrepant from the value found if the calculation involving the reported T^2 values is conducted. This slight difference occurs solely due to rounding in the reported T^2 values.)

Research Question 1

What is the relationship between different informants' ratings of anxiety, depression, or broader internalizing in youth with ASD?

Hypothesis 1a. The mean correlation effect size between youth self-report and parent report ratings of anxiety, depression, or broad internalizing in youth with ASD will be significant and yield a small to medium effect.

Mean correlations between youth self-report and parent report ratings were $r = 0.399$ ($p < 0.001$; $k = 25$; *95% CI* = 0.321, 0.471) for anxiety, $r = 0.412$ ($p < 0.001$; $k = 16$; *95% CI* = 0.313, 0.503) for depression, and $r = 0.430$ ($p < 0.001$; $k = 5$; *95% CI* = 0.242, 0.587) for broad internalizing. For all three constructs, the results were significant and observed values were consistent with a medium correlational effect size (Cohen, 1988; medium effect size $r \geq .30$ and $< .50$). Thus, hypothesis 1a was supported across all three constructs (i.e., anxiety, depression, and broad internalizing). See Tables 2-4 for detailed summaries of these results.

Hypothesis 1b. The mean correlation effect size between parent report and teacher report ratings of anxiety, depression, or broad internalizing in youth with ASD will be significant and yield a medium effect.

Mean correlations between parent and teacher ratings were $r = 0.273$ ($p < 0.001$; $k = 7$; *95% CI* = 0.185, 0.356) for anxiety, $r = 0.256$ ($p < 0.001$; $k = 7$; *95% CI* = 0.140, 0.366) for depression, and $r = 0.296$ ($p < 0.001$; $k = 9$; *95% CI* = 0.159, 0.422) for broad internalizing. For all three constructs, the results were significant and observed values were consistent with a small correlational effect size (Cohen, 1988; medium effect size $r \geq .10$ and $< .30$). Though all average correlation results were statistically significant, strictly speaking, hypothesis 1b was not supported because effect size estimates for anxiety, depression, and broad internalizing all fell short of Cohen's (1988) medium effect size standard of $r \geq .30$ and $< .50$. See Tables 5-7 for summaries of these results.

Research Question 1 Exploratory Analyses. A specific hypothesis was not made for the correlation between teacher report and self-report ratings of anxiety, depression, or broad internalizing in youth with ASD. However, findings regarding these correlations are reported here for the sake of completeness. The mean correlations between teacher and self-ratings were $r = 0.229$ ($p = 0.090$; $k = 2$; *95% CI* = -0.036, 0.464) for anxiety, $r = 0.342$ ($p = 0.097$; $k = 3$; *95% CI* = -0.064, 0.651) for depression, and $r = 0.316$ ($p = 0.255$; $k = 2$; *95% CI* = -0.233, 0.712) for broad internalizing. In all cases the 95% confidence interval around the mean correlation contained 0. Therefore, none of these results achieved statistical significance. (However, the small number of studies involved in each significance test should be considered, as low statistical power may have been a factor.) See Tables 8-10 for summaries of these results.

Research Question 2

Is the correlation between different informants' ratings of anxiety, depression, or broad internalizing in youth with ASD moderated by the general cognitive ability of the youth? See Table 11 for information on the studies that were included in these analyses.

Hypothesis 2a. The correlation between youth self-report ratings and parent ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the general cognitive ability of the youth, such that more advanced cognitive ability will be associated with better youth-parent agreement.

Initial tests of heterogeneity among effect size estimates for correlations between self-report and parent ratings were significant for anxiety (total true between-study variance: Q = 32.18, df = 12, p = 0.0013, T^2 = 0.0359, T = 0.1872, I^2 = 62.71%), depression (total between-study variance: Q = 26.25, df = 11, p = 0.0059, T^2 = 0.0362, T = 0.1904, I^2 = 58.10%), and broad internalizing (total between-study variance: Q = 9.02, df = 3, p = 0.0290, T^2 = 0.0368, T = 0.1919, I^2 = 66.75%). These results were suggestive of the presence of significant true variance in effect size estimates, which could be accounted for by one or more relevant moderator variables.

Despite the presence of potentially explainable effect size variation, results indicated that the correlation between youth self-report ratings and parent report ratings was not significantly moderated by the general cognitive ability, as measured by the youths' mean FSIQ score (continuous variable), for *anxiety* (test of the moderator meta-regression model: Q = 3.84, df = 1, p = 0.0502, k = 13, R^2_{Analog} = 0.30; regression equation: $Y_1 = 1.8993 - 0.0158X$ [Y_1 = predicted Fisher's z prime; X = mean FSIQ for a study]; test of residual variance: Q = 23.83, df = 11, p = 0.0135, T^2 = 0.0244, T = 0.1562, I^2 = 53.84%); *depression* (test of the moderator meta-regression

model: $Q = 0.68$, $df = 1$, $p = 0.4095$, $k = 12$, $R^2_{Analog} = 0.00$ [computed value: -0.15]; regression equation: $Y_1 = 1.4875 – 0.0104X$; test of residual variance: $Q = 25.76$, $df = 10$, $p = 0.0041$, $T^2 = 0.0415$, $T = 0.2038$, $I^2 = 61.18\%$); or *broad internalizing* (test of the moderator meta-regression model: $Q = 0.11$, $df = 1$, $p = 0.7369$, $k = 4$, $R^2_{Analog} = 0.00$ [computed value: -0.45]; regression equation: $Y_1 = 1.3566 – 0.0091X$; test of residual variance: $Q = 8.85$, $df = 2$, $p = 0.0120$, $T^2 = 0.0534$, $T = 0.2312$, $I^2 = 77.41\%$). Thus, hypothesis 2a was not supported across all three constructs (i.e., anxiety, depression, and broad internalizing) in the self vs. parent ratings context when cognitive ability was operationalized as the mean FSIQ score in each study.

Follow up analyses for 2a. Upon investigation of the scatter plot for FSIQ as a possible moderator of the correlation between parent and youth self-report of anxiety (see Figure 18 in Appendix G), it appeared that there were two distinct groups of studies: those with FSIQ mean values that fell generally below a mean FSIQ of 100 (Group 1; range: 87.78 to 94.98) and those with mean FSIQ values that generally fell above a mean FSIQ of 100 (Group 2; range = 101.66 to 110.14). Therefore, a follow-up, exploratory analysis was completed to separate these groups and the moderator analyses were re-run separately. The statistics for these groups were: Group 1 (total between-study variance*:* $Q = 6.40$, $df = 6$, $p = 0.3799$, $T^2 = 0.0019$, $T = 0.0433$, $I^2 = 6.24\%$; test of the moderator meta-regression model: $Q = 0.26$, $df = 1$, $p = 0.6115$, $k = 6$, $R^2_{Analog} = 0.00$ [computed value: -2.44]; regression equation: $Y_1 = 1.5628 – 0.0129X$; test of residual variance: $Q = 6.13$, $df = 5$, $p = 0.2940$, $T^2 = 0.0064$, $T = 0.0802$, $I^2 = 18.40\%$; Group 2 (total between-study variance*:* $Q = 10.50$, $df = 5$, $p = 0.0621$, $T^2 = 0.0182$, $T = 0.1348$, $I^2 = 52.40\%$; test of the moderator meta-regression model: $Q = 3.54$, $df = 1$, $p = 0.0599$, $k = 7$, $R^2_{Analog} = 0.67$; regression equation: $Y_1 = -2.9317 + 0.0376X$; test of residual variance: $Q = 5.21$, $df = 4$, $p = 0.2668$, $T^2 = 0.0059$, $T = 0.0768$, $I^2 = 23.17\%$). When the studies included in this moderator

analysis (FSIQ as a moderator between the correlation of parent and self-reported anxiety) were separated into two distinct groups, the obtained slopes of the two meta-regression equations differed in direction, However, in both cases the moderator variable failed to achieve statistical significance.

The scatterplot for FSIQ as a moderator between the correlation of parent vs. self-reported depression appeared to have an outlier present in the bottom right quadrant (see Figure 21 in Appendix G). Therefore, this analysis was rerun excluding the potential outlier. However, even without the outlier, the moderator analysis remained non-significant (total between-study variance*:* $Q = 19.71$, $df = 10$, $p = 0.0321$, $T^2 = 0.0241$, $T = 0.1551$, $I^2 = 49.26\%$; test of the moderator meta-regression model: $Q = 0.30$, $df = 1$, $p = 0.5854$, $k = 11$, $R^2_{Analog} = 0.00$ [computed value: -0.22]; regression equation: $Y_1 = 1.1062 - 0.0063X$; test of residual variance: $Q = 19.56$, $df = 9$, $p = 0.0208$, $T^2 = 0.0293$, $T = 0.1710$, $I^2 = 53.99\%$).

Hypothesis 2b. The correlation between youth self-report ratings and teacher ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the general cognitive ability of the youth, such that more advanced cognitive ability will be associated with better youth-teacher agreement.

There were not enough studies that reported correlations between self-report and teacher report ratings of anxiety ($k = 1$), depression ($k = 2$), or broad internalizing problems ($k = 1$) to complete moderator analyses of these variables.

Research Question 2 Exploratory Analyses. Although a specific hypothesis was not generated for general cognitive ability of the youth as a moderator of the correlation between parent and teacher ratings of anxiety, depression, and broad internalizing in youth with ASD, an

exploratory analysis was conducted for this rater-pair and reported here for purposes of completeness.

Initial tests of heterogeneity among effect size estimates for correlations between parent and teacher ratings were non-significant for *depression* (total between-study variance*: Q* = 0.27, *df* = 3, *p* = 0.9652, T^2 = 0.0000, *T* = 0.0000, I^2 = 0.00%) and *broad internalizing* (total between-study variance*: Q* = 1.90, *df* = 3, *p* = 0.5924, T^2 = 0.0000, *T* = 0.0000, I^2 = 0.00%). These results suggested insufficient true variability in effect size estimates, above and beyond variability expected due to sampling error, was available for moderator analyses involving this rater-pair. There were not enough studies for anxiety available to conduct moderator analyses for this rater-pair.

As expected, analyses indicated that the correlation between parent and teacher report ratings was not significantly moderated by general cognitive ability, as measured by the youths' mean FSIQ score, for *depression* (test of the moderator meta-regression model: *Q* = 0.12, *df* = 1, *p* = 0.7244, *k* = 4, R^2_{Analog} = 0.00; regression equation: Y_1 = 0.6112 – 0.0030*X*; test of residual variance: *Q* = 0.15, *df* = 2, *p* = 0.9652, T^2 = 0.0000, *T* = 0.0000, I^2 = 0.00%) or *broad internalizing* (test of the moderator meta-regression model: *Q* = 0.79, *df* =1 , *p* = 0.3735, *k* = 4, R^2_{Analog} = 0.00; regression equation: Y_1 = -0.5389 + 0.0076*X*; test of residual variance: *Q* = 1.11, *df* = 2, *p* = 0.5733, T^2 = 0.0000, *T* = 0.0000, I^2 = 0.00%). There were not enough studies available (*k* = 3) to complete moderator analyses for the correlation between parent and teacher ratings of anxiety (see Table 11).

Research Question 3

Are the correlations between self-report and other informants' ratings of anxiety, depression, or broad internalizing in youth with ASD moderated by the age of the youth? (See Table 12 for information concerning the studies that were included in these analyses.)

Hypothesis 3a. The correlation between self-report and parent ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the age of the youth with older age leading to a stronger positive correlation.

Initial tests of heterogeneity among effect size estimates for correlations between self-report and parent ratings were significant for *anxiety* (total between-study variance*: Q* = 50.21, *df* = 23, *p* = 0.0009, T^2 = 0.0262, *T* = 0.1617, I^2 = 54.19%), *depression* (total between-study variance*: Q* = 28.15, *df* = 14, *p* = 0.0136, T^2 = 0.0266, *T* = 0.1631, I^2 = 50.27%), and *broad internalizing* (total between-study variance*: Q* = 12.20, *df* = 4, *p* = 0.0159, T^2 = 0.0384, *T* = 0.1960, I^2 = 67.21%). These results were suggestive of the presence of significant true variance in effect size estimates, which could be accounted for by one or more relevant moderator variables.

Despite the presence of potentially explainable effect size variation, results indicated that the correlation between self-report and parent report ratings was not significantly moderated by the age of the youth, as measured as a continuous variable using the mean age in years of the youth sample in each study, for *anxiety* (test of the moderator meta-regression model: (*Q* = 1.23, *df* = 1, *p* = 0.2676, *k* = 24, R^2_{Analog} = 0.00 [computed value: -0.05]; regression equation: Y_1 = -0.0367 + 0.0364*X* [Y_1 = predicted Fisher's *z* prime; *X* = mean age for a study]; test of residual variance: *Q* = 49.23, *df* = 22, *p* = 0.0007, T^2 = 0.0275, *T* = 0.1658, I^2 = 55.31%); *depression* (test of the moderator meta-regression model: *Q* = 1.05, *df* = 1, *p* = 0.3050, *k* = 15, R^2_{Analog} = 0.16;

regression equation: Y_1 = -0.0030 + 0.0357X; test of residual variance: Q = 23.78, df = 13, p = 0.0332, T_2 = 0.0223, T = 0.1494, I_2 = 45.33%); or *broad internalizing* (test of the model: Q = 0.03, df = 1, p = 0.8673, k = 5, R^2_{Analog} = 0.00 [computed value: -0.42]; regression equation: Y_1 = 0.6563 – 0.0156X; test of residual variance: Q = 10.30, df = 3, p = 0.0162, T_2 = 0.0544, T = 0.2332, I_2 = 70.86%). Thus, hypothesis 3a was not supported for any of the three constructs (i.e., anxiety, depression, or broad internalizing) in the self-report vs. parent report context when age was operationalized as a continuous variable using mean values reported in the included studies.

Follow-up analyses for 3a. A review of the scatter plot for mean age as a possible moderator of the correlation between parent and youth self-report of depression, a potential outlier was observed. Therefore, this analysis was run again, this time without the outlier study, in order to examine whether this outlier had an impact on the statistical significance of the moderator analysis. When the outlier was removed from the analysis, the age moderator meta-regression model was found to be statistically significant (total between-study variance: Q = 21.71, df = 13, p = 0.0601, T_2 = 0.0169, T = 0.1300, I_2 = 40.11%; test of the moderator meta-regression model: Q = 5.28, df = 1, p = 0.0215, k = 14, R^2_{Analog} = 0.70; regression equation: Y_1 = -0.3408 + 0.0658X; test of residual variance: Q = 14.29, df = 13, p = 0.2826, T_2 = 0.0049, T = 0.0703, I_2 = 16.02%). Therefore, including this study in the moderator analysis may have hidden a relationship that may actually be occurring. Without the outlier included in the analysis, the results indicated that age did moderate the correlation between parent vs. self-reported depression (i.e., as age increased, so did the inter-rater correlation); thus, with the outlier removed, hypothesis 3a was supported for the depression construct.

Hypothesis 3b. The correlation between self-report and teacher ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the age of the youth with older age being associated with a stronger positive correlation.

As with hypothesis 2b, there were not enough studies that reported correlations between self-report and teacher report ratings of anxiety (k = 2), depression (k = 3), or broad internalizing problems (k = 2) to complete moderator analyses for hypothesis 3b (see Table 12).

Hypothesis 3c. The correlation between parent ratings and teacher ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the age of the youth with older age leading to a stronger positive correlation.

Initial tests of heterogeneity among effect size estimates for correlations between parent and teacher ratings were non-significant for *anxiety* (total between-study variance*:* Q = 4.40, df = 5, p = 0.4929, T^2 = 0.0000, T = 0.0000, I^2 = 0.00%) or *depression* (total between-study variance*:* Q = 8.10, df = 6, p = 0.2306, T^2 = 0.0067, T = 0.0819, I^2 = 25.96%). These results for anxiety and depression suggested insufficient true variability in effect size estimates, above and beyond variability expected due to sampling error, was available for moderator analyses involving this rater-pair. However, the heterogeneity of effect size estimates test was significant for *broad internalizing* (total between-study variance*:* Q = 20.51, df = 8, p = 0.0086, T^2 = 0.0274, T = 0.1655, I^2 = 61.00%), which was suggestive of the presence of significant true variance in effect size estimates that could be accounted for by one or more relevant moderator variables.

As expected, analyses indicated that the correlation between parent and teacher ratings was not significantly moderated by the mean age of the youth for *anxiety* (test of the moderator meta-regression model: Q = 0.08, df = 1, p = 0.7834, k = 6, R^2_{Analog} = 0.00; regression equation: Y_1 = 0.2056 + 0.0054X; test of residual variance: Q = 4.29, df = 4, p = 0.3684, T^2 =

0.0012, T = 0.0341, I^2 = 6.72%) or *depression* (test of the moderator meta-regression model: Q = 0.01, df = 1, p = 0.9073, k = 7, R^2_{Analog} = 0.00 [computed value: -0.57]; regression equation: Y_1 = 0.2477 + 0.0025X; test of residual variance: Q = 7.85, df = 5, p = 0.1648, T^2 = 0.0105, T = 0.1026, I^2 = 36.29%). However, it was also non-significant for *broad internalizing* (test of the moderator meta-regression model: Q = 1.51, df = 1, p = 0.2187, k = 9, R^2_{Analog} = 0.00 [computed value: -0.11]; regression equation: Y_1 = 0.5190 – 0.0275X; test of residual variance: Q = 19.09, df = 7, p = 0.0079, T^2 = 0.0305, T = 0.1747, I^2 = 63.33%). Thus, age of youth did not significantly moderate the correlation between parent and teacher ratings of anxiety, depression, or broad internalizing problems—and hypothesis 3c was not supported in the parent vs. teacher report context when age was operationalized as a continuous variable using mean values reported in the included studies.

Research Question 4

Is the correlation between self-report and parent ratings of anxiety, depression, or broad internalizing in youth with ASD moderated by the method of self-report administration (e.g., assessment read to child by researcher/clinician in the clinic, assessment read to the child by parent at home, assessment completed independently by the child in the clinic, etc.)?

Hypothesis 4. The correlation between self-report and parent ratings of anxiety, depression, or broad internalizing will be moderated by the method of self-report administration with the conditions "assessment read to child at home" and "assessment completed at home" being associated with a stronger positive correlation.

The categories for method of self-report administration that were included in these moderator analyses, and which reflected the range of categories available, were: (a) assessment read to child in clinic, (b) assessment completed in clinic, and (c) assessment completed at home.

No studies involved in these analyses reported that the self-report assessment was read to the child at home; therefore, this category was not represented.

The initial test of heterogeneity of effect size estimates for the correlations between self-report and parent report ratings of anxiety was significant (total between-study variance*:* Q = 26.39, df = 12, p = 0.0094, T^2 = 0.0237, T = 0.1541, I^2 = 54.53%). This result was suggestive of the presence of significant true variance in effect size estimates, which could be accounted for by one or more relevant moderator variables.

When run as a continuous variable, after dummy coding, the mean correlation between self-report and parent report of ***anxiety*** was significantly moderated by the method of self-report administration (test of the moderator meta-regression model: Q = 11.80, df = 2, p = 0.0027, k = 13, R^2_{Analog} = 0.82; regression equation: $Y_1 = 0.3643 - 0.0682X_1 + 0.2690X_2$ [Y_1 = predicted Fisher's z prime; X = self-report administration category); test of residual variance: Q = 12.14, df = 10, p = 0.2758, T^2 = 0.0044, T = 0.0660, I^2 = 17.63%) with the strongest positive correlation occurring in the "assessment completed at home" category (r = 0.559, k = 5). The "assessment read to child in clinic" condition yielded an r = 0.242 (k = 3) and the "assessment completed in clinic" condition yielded an r = 0.351 (k = 5).

In contrast, the correlation between self and parent ratings of ***depression*** was not significantly moderated by the method of self-report administration. The heterogeneity test suggested no substantive true variance in effect size estimates for any moderators to account for among the available studies (total between-study variance*:* Q = 3.54, df = 6, p = 0.7384, T^2 = 0.0000, T = 0.0000, I^2 = 0.00%). Thus, as expected, the test of the moderator meta-regression model (Q = 2.95, df = 2, p = 0.2282, k = 7, R^2_{Analog} = 0.00; regression equation: $Y_1 = 0.3070 + 0.1980X_1 + 0.2403X_2$) was not significant and neither was the test of residual variance (Q =

0.59, *df* = 4, *p* = 0.9645, T^2 = 0.0000, *T* = 0.0000, I^2 = 0.00%). Mean correlation values for each category were: *r* = 0.298 (*k* = 3) for assessment read to child in clinic, *r* = 0.477 (*k* = 2) for assessment completed in clinic, and *r* = 0.499 (*k* = 2) for assessment completed at home.

There were not enough studies (*k* = 4; see Table 13) that reported on method of self-report administration to run a moderator analysis for broad internalizing in the self-report vs. parent report ratings context. Thus, this was not examined statistically.

Overall, hypothesis 4 was supported for anxiety, not supported for depression, and could not be tested for broad internalizing. Thus, results for this hypothesis depended on the particular construct involved. (See Table 13 for more specific information on the studies that were included in these analyses.)

Follow-up analyses for hypothesis 4. In the scatterplot representing this moderator analysis within the parent vs. self-reported *anxiety* construct, there was considerably more variation in the category "assessment read to child in clinic" than in the other two categories ("assessment completed in clinic" and "assessment completed at home"). Additionally, there were only three studies that fell into this category. Due to the wide spread and small amount of data within the "assessment read to child in clinic" category and the fact that no studies could be specifically coded as falling cleanly into the "assessment read to child at home" category, the three available categories were collapsed to two new categories (i.e., assessment completed in clinic and assessment completed at home) to gain more statistical power for the comparison and acknowledge lack of clarity in most studies concerning whether the child had the items read to them or not in either setting. When the studies were regrouped into the two categories, the analysis maintained statistical significance and accounted for additional variance (test of the moderator meta-regression model: *Q* = 12.34, *df* = 1, *p* = 0.0004, *k* = 13, R^2_{Analog} = 0.89;

regression equation: $Y_1 = 0.3400 + 0.2927X$; test of residual variance: $Q = 12.40$, $df = 11$, $p = 0.3344$, $T^2 = 0.0026$, $T = 0.0510$, $I^2 = 11.28\%$). Specifically, the mean correlation in the "assessment completed in clinic" category was $r = 0.319$ ($k = 8$) and the mean correlation in the "assessment completed at home" category was $r = 0.559$ ($k = 5$), maintaining the previous pattern indicating that the correlation between parent vs. self-reported anxiety was generally higher when the assessment was completed at home.

For the sake of thoroughness, this approach of collapsing categories and re-running the analysis was also applied to the *depression* construct, even though the initial heterogeneity test indicated that there was no substantive variation present among the available studies for which a moderator could account. When the studies were regrouped into two categories, the analysis remained non-significant (test of the moderator meta-regression model: $Q = 1.84$, $df = 1$, $p = 0.1750$, $k = 7$, $R^2_{Analog} = 0.00$; regression equation: $Y_1 = 0.3676 + 0.1797X$; test of residual variance: $Q = 1.70$, $df = 5$, $p = 0.8885$, $T^2 = 0.0000$, $T = 0.0000$, $I^2 = 0.00\%$). Mean correlation values for each category were: $r = 0.361$ ($k = 5$) for assessment completed in clinic and $r = 0.501$ ($k = 2$) for assessment completed at home maintaining the pattern that correlation between parent vs. self-reported depression was typically higher when the assessment was completed at home.

Research Question 5

What are the mean differences between different informants' reports of anxiety, depression, or broad internalizing in youth with ASD?

Hypothesis 5a. When rating the behavior of youth with ASD, mean parent-rated anxiety, depression, or broad internalizing scores will be significantly higher than mean youth self-report ratings.

In these analyses, a positive effect size *g* reflects parent ratings > youth self-report ratings and a negative *g* reflects parent ratings < youth self-report ratings. The standardized mean difference between *parent-rated and self-rated anxiety* was statistically significant—with parent ratings yielding a higher mean, reflecting higher anxiety or more anxiety symptoms, than self-ratings (g = 0.220 [p < 0.001; k = 46; *95% CI* = 0.102, 0.337]). This standardized mean difference result was consistent with a small effect (see Cohen, 1988 under effect size *d*, which is interpreted on the same metric as *g*; small effect is ≥ .20 and < .50). *Parent-rated and self-rated mean values for depression* were also significantly different—with parents endorsing higher depression or more depression symptoms than the youths themselves (g = 0.788 [p < 0.001; k = 18; *95% CI* = 0.501, 1.074]). This standardized mean difference result was consistent with a medium effect (see Cohen, 1988 under effect size *d*, which is interpreted on the same metric as *g*; medium effect is ≥ .50 and < .80). However, the observed value was very close to (0.012 points under) the minimum standard for a large effect (i.e., ≥ .80), with the 95% confidence interval overlapping considerably with both the medium and large effect size ranges.

In contrast, the standardized mean difference between *parent and self-ratings of youth broad internalizing problems* was not statistically significant (g = 0.090 [p = 0.341; k = 4; *95% CI* = -0.095, 0.276]). In this case, the observed standardized mean difference result was negligible, as it did not meet the minimum standard for a small effect (see Cohen, 1988 under effect size *d*, which is interpreted on the same metric as *g*; small effect is ≥ .20 and < .50) and was close to zero.

Overall, hypothesis 5a was supported for both anxiety and depression—with mean parent ratings significantly exceeding youth self-report ratings. However, hypothesis 5a was not supported, based on an analysis of four studies, for broad internalizing. (See Tables 14-16 for

summaries of these results. Also, see the publication bias subsection at the end of the results section for additional qualifying information regarding the anxiety and depression results in the parent vs. youth self-report context.)

Hypothesis 5b. When rating the behavior of youth with ASD, mean teacher-rated anxiety, depression, or broad internalizing scores will be significantly higher than mean youth self-report ratings.

It is important to note that there was considerable variability in the effect size estimates, for teacher vs. youth self-report ratings, across studies for all three constructs (see Tables 17-19). Overall, the teacher-rated and self-rated mean values for *anxiety* (g = 0.295 [p = 0.417; k = 6; *95% CI* = -0.417, 1.006]), *depression* (g = 0.670 [p = 0.097; k = 4; *95% CI* = -0.121, 1.461]), and *broad internalizing* (g = -0.033 [p = 0.930; k = 2; *95% CI* = -0.770, 0.704]) scores did not statistically differ. In terms of observed effect size, the standardized mean difference estimates were negligible (≥ 0 and < .20) for *broad internalizing*, small (≥ .20 and < .50) for *anxiety*, and medium (≥ .50 and < .80) for *depression*. Overall, hypothesis 5b was not supported for any of the three constructs (see Tables 17-19 for details of the studies involved).

Hypothesis 5c. When rating the behavior of youth with ASD, mean teacher-rated anxiety, depression, or broad internalizing scores will not differ substantially from mean parent ratings.

Mean teacher ratings did not differ significantly from parent ratings for *depression* (g = 0.176 [p = 0.349; k = 9; *95% CI* = -0.192, 0.545]), but did differ significantly for *anxiety* (g = 0.156 [p = 0.002; k = 11; *95% CI* = 0.058, 0.254]) and *broad internalizing* (g = 0.153 [p = 0.041; k = 14; *95% CI* = 0.006, 0.299]). However, the obtained standardized mean difference effect size estimates were negligible (i.e., ≥ 0 and < .20) for all three comparisons. Therefore,

hypothesis 5c was supported, as teacher vs. parent ratings did not differ substantially for anxiety, depression, or broad internalizing. (See Tables 20-22 for details of the studies involved.)

Research Question 6

Are the mean differences between informant's ratings of anxiety, depression, or broad internalizing in youth with ASD moderated by the general cognitive ability of the youth? (See Table 23 for information on the studies that were included in these analyses.)

Hypothesis 6a. Mean differences between youth self-report ratings and parent report ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by youth cognitive ability with more advanced cognitive ability being associated with a smaller mean difference.

Initial tests of heterogeneity among effect size estimates for mean differences between parent ratings and youth self-report ratings were significant for *anxiety* (total between-study variance*:* $Q = 75.15$, $df = 24$, $p < 0.001$, $T^2 = 0.0547$, $T = 0.2339$, $I^2 = 68.06\%$) and *depression* (total between-study variance*:* $Q = 56.85$, $df = 11$, $p = < 0.001$, $T^2 = 0.2165$, $T = 0.4653$, $I^2 = 80.65\%$). These results suggest the presence of a significant true variance in effect size estimates, which may be accounted for by relevant moderator variables. In the case of broad internalizing ratings, there were insufficient data available to complete moderator analyses for cognitive ability.

Although the heterogeneity tests indicated the presence of potentially explainable effect size variation, results suggested that the mean differences between youth self-report ratings and parent report ratings were not significantly moderated by the general cognitive ability of the youth (measured as a continuous variable using the youths' mean FSIQ score) for *anxiety* (test of the moderator meta-regression model: $Q = 0.13$, $df = 1$, $p = 0.722$, $k = 25$, $R^2_{Analog} = 0.00$

[computed value: -0.06]; regression equation: Y_1 = -0.0707 + 0.0029(X); test of residual variance: Q = 72.89, df = 23, p = < 0.001, T_2 = 0.0578, T = 0.2404, I_2 = 68.45%) or d*epression* (test of the moderator meta-regression model: Q = 2.03, df = 1, p = 0.154, k = 12, R^2_{Analog} = 0.31; regression equation: Y_1 = -1.5090 + 0.0224(X); test of residual variance: Q = 37.96, df = 10, p = 0.0000, T_2 = 0.1501, T = 0.3874, I_2 = 73.66%). Overall, mean differences in ratings of anxiety and depression do not appear to be moderated by youth cognitive ability in the youth self-report vs. parent report ratings context. Thus, hypothesis 6a was not supported.

Hypothesis 6b. Mean differences between youth self-report ratings and teacher report ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by youth cognitive ability with more advanced cognitive ability being associated with a smaller mean difference.

The test of heterogeneity among effect size estimates for mean differences between youth self-report ratings and teacher report ratings of anxiety were significant (total between-study variance*:* Q = 21.08, df = 3, p < 0.001, T_2 = 0.4731, T = 0.6878, I_2 = 85.77%) indicating a presence of significant true variance in effect size estimates accounted for by one or more moderator variables. Unfortunately, there were not enough studies that clearly described the cognitive ability of the youth sample to complete this moderator analysis for depression or broad internalizing problems.

Despite the potential explainable effect size variation indicated in the initial test of heterogeneity, results suggested that the mean differences between youth self-report ratings and teacher report ratings of *anxiety* were not significantly moderated by the cognitive ability of the youth as measured using the youths' mean FSIQ score (test of the moderator meta-regression model: Q = 0.00, df = 1, p = 0.971, k = 4, R^2_{Analog} = 0.00 [computed value: -0.94]; regression

equation: $Y_1 = 0.3765 – 0.0016(X)$; test of residual variance: $Q = 20.62$, $df = 2$, $p < 0.001$, $T_2 = 0.9196$, $T = 0.9590$, $I_2 = 90.30\%$). Therefore, hypothesis 6b was not supported for anxiety and could not be tested statistically for depression or broad internalizing.

Research Question 6 Exploratory Analyses. A specific hypothesis was not generated for youth cognitive ability as a moderator of the mean differences between parent report and teacher report ratings of anxiety, depression, or broad internalizing. However, this was investigated as an exploratory analysis for the sake of completeness.

Initial heterogeneity tests among effect sizes for mean differences between parent and teacher ratings were significant for *anxiety* (total between-study variance*:* $Q = 63.40$, $df = 6$, $p < 0.001$, $T_2 = 0.2202$, $T = 0.4692$, $I_2 = 90.54\%$), *depression* (total between-study variance*:* $Q = 75.95$, $df = 5$, $p < 0.001$, $T_2 = 0.2960$, $T = 0.5441$, $I_2 = 93.42\%$), and *broad internalizing* (total between-study variance*:* $Q = 31.89$, $df = 6$, $p < 0.001$, $T_2 = 0.0742$, $T = 0.2723$, $I_2 = 81.19\%$), which suggests that there is the presence of true variance that could be accounted for by moderator variables.

However, results indicated that standardized mean differences were not significantly moderated by cognitive ability of the youth (mean FSIQ value) for *anxiety* (test of the moderator meta-regression model: $Q = 0.08$, $df = 1$, $p = 0.7835$, $k = 7$, $R^2_{Analog} = 0.07$; regression equation: $Y_1 = 0.5093 – 0.0034(X)$; test of residual variance: $Q = 39.38$, $df = 5$, $p < 0.001$, $T_2 = 0.2045$, $T = 0.4522$, $I_2 = 87.30\%$), *depression* (test of the moderator meta-regression model: $Q = 1.64$, $df = 1$, $p = 0.1999$, $k = 6$, $R^2_{Analog} = 0.46$; regression equation: $Y_1 = 1.6636 – 0.0145(X)$; test of residual variance: $Q = 29.63$, $df = 4$, $p < 0.001$, $T_2 = 0.1609$, $T = 0.4011$, $I_2 = 86.50\%$), or *broad internalizing* (test of the moderator meta-regression model: $Q = 0.08$, $df = 1$, $p = 0.7824$, k

= 7, R^2_{Analog} = 0.00 [computed value: -0.17]; regression equation: $Y_1 = 0.4846 – 0.0030(X)$; test of residual variance: $Q = 29.04$, $df = 5$, $p < 0.001$, $T^2 = 0.0869$, $T = 0.2947$, $I^2 = 82.78\%$).

Research Question 7

Are the mean differences between informant's reports of anxiety, depression, or broad internalizing in youth with ASD moderated by the age of the child? (See Table 24 for information on the studies that were included in these analyses.)

Hypothesis 7a. Mean differences between youth self-report and parent report ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by youth age with older age being associated with a smaller mean difference.

Initial tests of heterogeneity among effect size estimates for mean differences between youth self-report ratings and parent report ratings were significant for *anxiety* (total between-study variance: $Q = 212.90$, $df = 41$, $p < 0.001$, $T^2 = 0.1082$, $T = 0.3289$, $I^2 = 80.74\%$) and *depression* (total between-study variance: $Q = 139.25$, $df = 15$, $p < 0.001$, $T^2 = 0.3461$, $T = 0.5883$, $I^2 = 89.23\%$). These results suggest the presence of significant true variance in effect size estimates that may be accounted for my relevant moderator variables. In the case of broad internalizing ratings, there were not enough studies available for the moderator analysis regarding youth age on self- vs. parent ratings of broad internalizing problems ($k = 3$; see Table 24).

Although the tests of heterogeneity indicated that a moderator variable may be present, the mean differences between youth self-report ratings and parent report ratings were not significantly moderated by youth age (measured as a continuous variable using the mean age in years reported in each study) for *anxiety* (test of the moderator meta-regression model: $Q = 0.04$, $df = 1$, $p = 0.847$, $k = 42$, $R^2_{Analog} = 0.00$ [computed value: -0.04]; regression

equation: $Y_1 = 0.2861 – 0.0069(X)$; test of residual variance: $Q = 212.89$, $df = 40$, $p < 0.001$, $T_2 = 0.1121$, $T = 0.3349$, $I_2 = 81.21\%$) or ***depression*** (test of the moderator meta-regression model: $Q = 1.03$, $df = 1$, $p = 0.310$, $k = 16$, $R_{2Analog} = 0.00$ [computed value: -0.07]; regression equation: $Y_1 = 1.8544 – 0.0878(X)$; test of residual variance: $Q = 133.92$, $df = 14$, $p < 0.001$, $T_2 = 0.3699$, $T = 0.6082$, $I_2 = 89.55\%$). Overall, mean differences in ratings of anxiety and depression do not appear to be moderated by youth age in the youth self-report vs. parent report ratings context. Thus, hypothesis 7a was not supported for anxiety and depression, while a meaningful test could not be conducted for broad internalizing in this context.

Hypothesis 7b. Mean differences between youth self-report and teacher report ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by youth age with older age associated with smaller mean differences.

Based on number of studies available, the only moderator analysis that could be completed to investigate age of youth as a moderator was for youth self-report and teacher report ratings of anxiety. Initial tests of heterogeneity among effect size estimates within this rater-pair and construct were significant (total between-study variance*:* $Q = 30.03$, $df = 4$, $p < 0.001$, $T_2 = 0.8884$, $T = 0.9425$, $I_2 = 86.68\%$). However, despite the presence of potentially explainable effect size variation, results indicated that the standardized mean differences between youth self-ratings and teacher ratings of ***anxiety*** were not significantly moderated by youth age (test of the moderator meta-regression model: $Q = 0.01$, $df = 1$, $p = 0.932$, $k = 5$, $R_{2Analog} = 0.00$ [computed value: -0.64]; regression equation: $Y_1 = 0.7842 – 0.0306(X)$; test of residual variance: $Q = 27.32$, $df = 3$, $p < 0.001$, $T_2 = 1.4528$, $T = 1.2053$, $I_2 = 89.02\%$). Thus, hypothesis 7b was not supported for anxiety, while meaningful hypothesis tests were not possible for depression and broad internalizing.

Hypothesis 7c. Mean differences between parent ratings and teacher ratings of anxiety, depression, or broad internalizing in youth with ASD will be moderated by youth age with older age associated with a smaller mean difference.

Based on tests of heterogeneity among effect size estimates for mean differences between parent and teacher ratings, *anxiety* (total between-study variance*: Q* = 81.83, *df* = 8, *p* < 0.001, T^2 = 0.2704, *T* = 0.5200, I^2 = 90.22%), *depression* (total between-study variance*: Q* = 81.96, *df* = 6, *p* < 0.001, T^2 = 0.2903, *T* = 0.5388, I^2 = 92.68%), and *broad internalizing* (total between-study variance*: Q* = 52.47, *df* = 11, *p* < 0.001, T^2 = 0.0530, *T* = 0.2302, I^2 = 79.04%) were significant for the presence of true variance accounted for my relevant moderators.

Despite this significance, results showed that mean differences between parent and teacher ratings were not significantly moderated by youth age for *anxiety* (test of the moderator meta-regression model: *Q* = 0.60, *df* = 1, *p* = 0.438, *k* = 9, R^2_{Analog} = 0.00 [computed value: -0.20]; regression equation: Y_1 = -0.4426 + 0.0466(*X*); test of residual variance: *Q* = 79.25, *df* = 7, *p* < 0.001, T^2 = 0.3254, *T* = 0.5704, I^2 = 91.17%), *depression* (test of the moderator meta-regression model: *Q* = 0.00, *df* = 1, *p* = 0.968, *k* = 7, R^2_{Analog} = 0.00 [computed value: -0.11]; regression equation: Y_1 = 0.2631 – 0.0025(*X*); test of residual variance: *Q* = 74.79, *df* = 5, *p* < 0.001, T^2 = 0.3212, *T* = 0.5668, I^2 = 93.31%), or *broad internalizing* (test of the moderator meta-regression model: *Q* = 1.08, *df* = 1, *p* = 0.298, *k* = 12, R^2_{Analog} = 0.00 [computed value: -0.17]; regression equation: Y_1 = -0.0759 + 0.0293(*X*); test of residual variance: *Q* = 52.42, *df* = 10, *p* < 0.001, T^2 = 0.0620, *T* = 0.2302, I^2 = 80.92%). Thus, age of youth—when measured as a continuous variable using the mean age in years reported in each study—did not significantly moderate the standardized mean differences between parent and teacher ratings of anxiety,

depression, or broad internalizing—and, therefore, hypothesis 7c was not supported in the parent vs. teacher report context.

Research Question 8

Are the mean differences between informant's ratings of anxiety, depression, or broad internalizing in youth with ASD moderated by the method of self-report administration?

Hypothesis 8. Differences in mean scores between youth self-report and parent report of anxiety, depression, or broad internalizing in youth with ASD will be moderated by the method of self-report administration with the conditions "assessment read to child at home" and "assessment completed at home" leading to a smaller mean difference.

Moderator analyses regarding method of self-report administration and the differences in mean scores between youth and parent ratings of anxiety and depression were conducted; however, there were not enough studies available to investigate whether the method of self-report administration moderated the standardized difference in mean scores between youth self-report and parent report of broad internalizing problems ($k = 3$; see Table 25). For the analysis of self-report administration within the ***anxiety*** construct, the conditions "assessment read to child in clinic" ($g = 0.468$), "assessment completed in clinic" ($g = 0.367$), and "assessment completed at home" ($g = 0.131$) were represented. For the analysis of self-report administration within the ***depression*** construct all four conditions were represented ("assessment read to child in clinic" [$g = 0.980$], "assessment read to child at home" [$g = 1.001$], "assessment completed in the clinic" [$g = 1.023$], and "assessment completed at home" [$g = 0.253$].

The initial test of heterogeneity of effect size estimates for the mean differences between self-report and parent report ratings of ***anxiety*** (total between-study variance*:* $Q = 134.67$, $df = 20$, $p < 0.001$, $T^2 = 0.1261$, $T = 0.3551$, $I^2 = 85.15\%$) and ***depression*** (total between-study

variance*:* $Q = 120.52$, $df = 11$, $p < 0.001$, $T^2 = 0.3993$, $T = 0.6319$, $I^2 = 90.87\%$) were significant. This could mean that there is significant true variance in effect size estimates accounted for by relevant moderator variables.

Despite the results from the initial test of heterogeneity, the standardized mean differences between youth and parent ratings of *anxiety* (test of the moderator meta-regression model: $Q = 2.18$, $df = 2$, $p = 0.3368$, $k = 21$, $R^2_{Analog} = 0.06$; regression equation: $Y_1 = 0.4815 – 0.2137(X_1) – 0.3418(X_2)$; test of residual variance: $Q = 114.50$, $df = 18$, $p < 0.000$, $T^2 = 0.1181$, $T = 0.3437$, $I^2 = 84.28\%$) and *depression* (test of the moderator meta-regression model: $Q = 4.19$, $df = 3$, $p = 0.242$, $k = 12$, $R^2_{Analog} = 0.33$; regression equation: $Y_1 = 0.9741 + 0.0270(X_1) + 0.0403(X_2) – 07.228(X_3)$; test of residual variance: $Q = 62.78$, $df = 8$, $p < 0.000$, $T^2 = 0.2677$, $T = 0.5174$, $I^2 = 87.26\%$) were not significantly moderated by the method of self-report administration. Thus, hypothesis 8 was not supported for either anxiety or depression, and could not be tested for broad internalizing. See Table 25 for information on the studies that were included in these analyses.

Follow-up analyses for hypothesis 8. Consistent with the process described in the follow-up analysis section under hypothesis 4, method of self-report administration categories were also collapsed and re-run for the moderator analyses regarding method of self-report administration and the differences in mean scores between parent vs. self-report of anxiety and depression.

Within the constructs of parent vs. self-reported *anxiety* (test of the moderator meta-regression model: $Q = 1.40$, $df = 1$, $p = 0.2370$, $k = 21$, $R^2_{Analog} = 0.00$ [computed value: -0.01]; regression equation: $Y_1 = 0.3519 – 0.2108(X)$; test of residual variance: $Q = 127.31$, $df = 19$, $p < 0.001$, $T^2 = 0.1269$, $T = 0.3563$, $I^2 = 85.08\%$) and *depression* (test of the moderator meta-

regression model: $Q = 3.04$, $df = 1$, $p = 0.0812$, $k = 12$, $R^2_{Analog} = 0.35$; regression equation: $Y_1 = 0.9978 – 0.5939(X)$; test of residual variance: $Q = 70.97$, $df = 10$, $p < 0.001$, $T^2 = 0.2603$, $T = 0.5102$, $I^2 = 85.91\%$), the moderator analyses remained non-significant. However, both patterns were consistent with smaller mean differences within the "assessment completed at home" category (*parent vs. self-reported anxiety:* "assessment completed in clinic" $g = 0.384$ [$p = 0.015$, $k = 12$], "assessment completed at home" $g = 0.131$ [$p = 0.214$, $k = 9$]; *parent vs. self-reported depression:* "assessment completed in clinic" $g = 0.984$ [$p < 0.001$, $k = 8$], "assessment completed at home" $g = 0.412$ [$p = 0.179$, $k = 4$].

Investigation of Correlation Coefficient Type as a Potential Moderator

As described above, each correlation coefficient (i.e., Pearson's *r*, Spearman's *rho,* and ICC) was treated as Pearson's *r* in order to maximize the number of studies that were included in the correlation analysis of this meta-analysis. To investigate the impact of this decision, exploratory analyses were conducted wherein the correlation coefficient type reported in each study was evaluated as a potential moderator for parent vs. self and parent vs. teacher ratings. (There were insufficient studies available to complete this moderator analysis for any construct within the teacher vs. self-rated pair [anxiety: $k = 2$; depression: $k = 3$; internalizing: $k = 2$; see Tables 8–10 or Table 26].) Pearson's *r* was the most commonly reported correlation coefficient for parent vs. self-rated anxiety and depression, and for parent vs. teacher-rated anxiety, depression, and broad internalizing. For the five studies included in the correlational analysis for parent vs. self-rated broad internalizing, Pearson's *r* and ICC were both reported twice and Spearman's *rho* was reported once. (See Tables 2-7 or Table 26 for correlation type reported within each study).

Initial tests of heterogeneity among effect size estimates for correlation between self-report and parent report were significant for *anxiety* (total between-study variance*: Q* = 51.18, *df* = 24, *p* = 0.0010 , T^2 = 0.0257, *T* = 0.1603, I^2 = 53.11%, *depression* (total between-study variance*: Q* = 28.16, *df* = 15, *p* = 0.0206, T^2 = 0.0239, *T* = 0.1547, I^2 = 46.73%), and *broad internalizing* (total between-study variance*: Q* = 12.20, *df* = 4, *p* = 0.0159, T^2 = 0.0383, *T* = 0.1960, I^2 = 67.21%). These results suggest the presence of true variance in effect size estimates, which can be accounted for by relevant moderators.

Despite the presence of potentially explainable effect size variation, results showed that the correlation between parent and self-ratings of *anxiety* (test of the moderator meta-regression model: *Q* = 4.58, *df* = 2, *p* = 0.1011, *k* = 25, R^2 = 0.09; regression equation: Y_1 = 0.3825 – 0.0922(X_1) + 0.2117(X_2); test of residual variance: *Q* = 51.18, *df* = 22, *p* = 0.0010, T^2 = 0.0257, *T* = 0.1603, I^2 = 53.11%), *depression* (test of the moderator meta-regression model: *Q* = 2.05, *df* = 2, *p* = 0.3595, *k* = 16, R^2 = 0.00 [computed value: -0.06]; regression equation: Y_1 = 0.4172 – 0.1186(X_1) + 0.22829(X_2); test of residual variance: *Q* = 24.90, *df* = 13, *p* = 0.0238, T^2 = 0.0254, *T* = 1593, I^2 = 47.79%), and *broad internalizing* (test of the moderator meta-regression model: *Q* = 0.65, *df* = 2, *p* = 0.7216, *k* = 5, R^2 = 0.00 [computed value: -0.35]; regression equation: Y_1 = 0.5444 – 0.2676(X_1) – 0.0988(X_2); test of residual variance: *Q* = 6.92, *df* = 2, *p* = 0.0315, T^2 = 0.0520, *T* = 0.2281, I^2 = 71.09%) were not statistically significant.

Similarly, the tests of heterogeneity among effect size estimates for correlations between parent and teacher ratings of *anxiety* (total between-study variance*: Q* = 6.39, *df* = 6, *p* = 0.3807, T^2 = 0.0010, *T* = 0.0320, I^2 = 6.14%), *depression* (total between-study variance*: Q* = 8.10, *df* = 6, *p* = 0.2306, T^2 = 0.0067, *T* = 0.0819, I^2 = 25.96%), and *broad internalizing* (total

between-study variance*:* Q = 20.51, df = 8, p =0.0086, T^2 = 0.0274, T = 0.1655, I^2 = 61.00%) were significant.

Although the tests of heterogeneity suggest possible true variance that could be accounted for my moderators, the correlation type moderation analyses did not achieve statistical significance for the correlation between parent and teacher ratings of *anxiety* (test of the moderator meta-regression model: Q = 1.66, df = 2, p = 0.4350, k = 7, R^2 = 0.00 [computed value: -0.38]; regression equation: Y_1 = 0.2339 + 0.1363(X_1) + 0.1163(X_2); test of residual variance: Q = 4.35, df = 4, p = 0.3604, T^2 = 0.0014, T = 0.0376, I^2 = 8.09%), *depression* (test of the moderator meta-regression model: Q = 0.50, df = 2, p = 0.7802, k = 7, R^2 = 0.00 [computed value: -0.71]; regression equation: Y_1 = 0.2410 + 0.0796(X_1) + 0.1143(X_2); test of residual variance: Q = 6.80, df = 4, p = 0.1470, T^2 = 0.0115, T = 0.1072, I^2 = 41.16%) or *broad internalizing* (test of the moderator meta-regression model: Q = 1.44, df = 2, p = 0.4867, k = 9, R^2 = 0.00 [computed value: -0.19]; regression equation: Y_1 = 0.3573 – 0.3073(X_1) – 0.1582(X_2); test of residual variance: Q = 18.91, df = 6, p = 0.0043, T^2 = 0.0327, T = 0.1807, I^2 = 68.27%). Therefore, for parent vs. self-ratings and parent vs. teacher ratings, treating all correlation estimates as if they were Pearson's *r* did not appear to significantly impact the average correlation estimates between rater-pairs. Again, there were too few studies (k = 2-3) to test this hypothesis in the teacher report vs. self-rating context (see Table 26).

Publication Bias

Three methods were employed for examining possible publication bias. These methods were the Rosenthal *fail-safe N*, visual inspection of funnel plots, and implementation of the Trim and Fill method (Duval & Tweedie, 2000) when funnel plots suggested evidence of possible bias. No bias analyses were conducted for three of the above effect size results (i.e., correlation

between teacher vs. self-reported anxiety, correlation between teacher vs. self-reported broad internalizing, and the standardized mean difference between teacher vs. self-reported broad internalizing), because fewer than 3 studies were involved in each. For the fifteen overall effect size results involving 3 or more studies, the relevant *fail-safe N* results are reported beneath each of the 15 funnel plots (see Figures 3 – 17 in Appendix F).

Rosenthal's *fail-safe N* yielded a result of zero for three of the 15 overall effect size estimates. These *fail-safe N* results were for: (a) mean differences between teacher and self-reported anxiety ($z = 1.69$; $p = 0.09$; *Fail-safe N* = 0), (b) mean differences between parent and self-reported broad internalizing problems ($z = 1.37$; $p = 0.17$; *Fail-safe N* = 0), and (c) mean differences between parent and teacher reported anxiety ($z = 0.87$; $p = 0.38$; Fail-safe $N = 0$). However, close inspection of each of these results indicated that the observed overall mean effect size result were either non-significant or negligible in each case. Given that Rosenthal's *Fail-safe N* is an estimate of how many theoretically missing or unpublished studies with zero effect would be needed to overturn a statistically significant effect size result, it was irrelevant in these cases because the obtained overall mean effect size estimates were already non-significant. *Fail-safe N* results for the other 12 overall effect size estimates were generally not of concern.

Additionally, if all relevant studies have been published in some form and a meta-analysis has captured all of the relevant studies for analysis, the funnel plot will tend to look symmetrical (i.e., effect sizes for the studies are dispersed equally on either side of the overall effect). However, if the funnel plot is asymmetrical (especially when more of the smaller studies tend to appear on one side of the mean overall effect), there is concern that studies that would theoretically fall on the opposite side of the mean effect are missing from the analysis (Borenstein et al., 2014). In theory, this pattern can occur when smaller studies with non-

significant results are less likely to be published. Duval and Tweedie's Trim and Fill method (Duval & Tweedie, 2000) was developed to estimate and attempt to correct for what is missing. Specifically, Borenstein et al., (2014) indicated that the Trim and Fill process initially trims the asymmetric studies from one side to locate a more unbiased effect size estimate and then imputes more balanced estimates for missing individual studies on both sides of the mean overall effect to better estimate variability.

Based on visual inspection of the computed funnel plots, two plots were selected for further investigation (see Figures 10 and 11 in Appendix F) due to the asymmetry of the plots. For the funnel plot reflecting standardized mean differences between parent vs. self-reported anxiety (Figure 10), the Trim and Fill method suggested that there were 9 potential studies missing from the left side of the mean effect (indicating that small studies may have been more likely to be published when the results revealed a larger positive effect size). Under the random effects model, the original point estimate and 95% confidence interval for the combined studies was 0.21963 (0.10204, 0.33722). Using Trim and Fill, the bias-adjusted point estimate and 95% confidence interval was 0.05815 (-0.07045, 0.18674). Because the less biased estimate (i.e., 0.05815) suggests a negligible effect, it is possible that smaller published studies were more likely to report significant differences in the report of anxiety wherein parents gave higher ratings than the youth self-ratings. This less biased estimate was non-significant and close to 0, which suggests that, on average, parent vs. self-report means for anxiety may be more similar—yielding close to no mean difference between the two rater types). However, even if the original obtained estimate (0.21963) is accurate, the difference suggesting parents tend to rate anxiety higher than the child self-report was small and may not be considered substantive.

The Trim and Fill method was also utilized for investigating the studies included in the analysis for mean differences between parent vs. self-reported depression (see Figure 11 in Appendix F). Again, this showed that 9 potential studies were missing from the left side of the mean effect. Under the random effects model, the original observed point estimate and 95% confidence interval for the combined studies was 0.78760 (0.50086, 1.07433). Using Trim and Fill, the imputed point estimate and 95% confidence interval was 0.24430 (-0.05600, 0.54461). After employing this method, the point estimate went from a medium effect to a small effect. This suggests that smaller published studies were more likely to report significant and substantive differences between the raters of depression with, on average, parents providing higher ratings than the youth themselves. Critically, the bias-adjusted standardized mean difference estimate (g = 0.24430) was substantially smaller than the unadjusted, observed estimate (g = 0.78760)—with the 95% confidence interval around the bias-adjusted estimate containing 0 and falling short of the threshold for statistical significance. Regardless of what the true standardized mean difference is between parent and youth self-report for depression, these findings suggest it is likely considerably smaller than the observed point result—a pattern which is consistent with effect size inflation resulting from possible publication bias.

CHAPTER V

DISCUSSION

Brief Study Rationale and Overview

Internalizing problems such as anxiety and depression are among the most common psychiatric comorbidities within ASD (Davidsson et al., 2017; DSM-5, 2013; Lopata et al., 2010; Park et al., 2013; Strang et al., 2012; van Steensel & Heeman, 2017). These internalizing problems can result in considerable negative impact on the lives of children and adolescents (Bellini, 2004; Kim et al., 2011; Matson & Williams, 2014; Michael & Merrell, 1998). For example, internalizing problems can adversely affect mental health, physical health, self-esteem, social competence, attention/concentration, academic performance, and quality of life (Huberty, 2014; Kerns & Kendall, 2014; Michael & Merrell, 1998).

Efforts to identify internalizing problems in youth are critical, especially in populations of youth that are more at-risk for these problems. Therefore, understanding the reliability and validity of the various tools and strategies used for the assessment of internalizing problems is essential. Given the emphasis on use of multiple sources and multiple methods, as part of a best practice assessment strategy (Taylor et al., 2018), clinicians need research-based guidance concerning such issues as the level of agreement across sources and methods, conditions under which agreement may vary, conditions under which one source or method may be more informative than another, and how to effectively synthesize data from across sources and methods. Use of behavior rating scales is a major method utilized for screening or as a part of a more comprehensive assessment for internalizing issues, in ASD and more generally (Merrell et al., 2002), which allows for ratings from multiple sources (e.g., self, parent/caregiver, teacher/daycare provider, etc.). Reports on the level of agreement across these rater sources vary

considerably within the larger research literature concerning internalizing issues in youth with ASD (Barnhill et al., 2000; Blakeley-Smith et al., 2012; Chow, 2008; Hurtig et al., 2009; Kaat & Lecavalier, 2015; Lopata et al., 2010; Magiati et al., 2014; Ooi et al., 2016; Taylor et al., 2018; Volker et al., 2010; White et al., 2011) and concerns exists regarding the level of insight that youth with ASD may have for the self-evaluation of emotions and other internal states (Baron-Cohen, 2002; Lopata et al., 2010; Mazefsky et al., 2011). This situation makes general interpretations difficult in the absence of a more comprehensive approach to summarizing and synthesizing the inter-rater or cross-informant ratings of anxiety, depression, and broader internalizing among youth with ASD. To date, no meta-analysis has been conducted for this purpose.

This study investigated and summarized the level of agreement across different combinations of rater-pairs assessing internalizing problems in youth with ASD. Analyses focused on both inter-rater correlations and cross-rater standardized mean differences in order to gain a more comprehensive understanding of inter-rater agreement for this population. This meta-analysis also examined the impact of potential moderator variables such as youth cognitive ability, youth age, method of self-report administration, and type of correlation coefficient reported on appropriate effect size estimates.

Correlational and Mean Difference Effect Size Analyses Across Constructs and Rater-Pairs

The average correlation effect size estimates and mean difference effect size estimates for each construct (i.e., anxiety, depression, and broad internalizing) within each rater-pair (i.e., parent vs. self, parent vs. teacher, and self vs. teacher) will be examined in relation to prior salient findings cited in the literature review and in comparison to findings from prior meta-analyses involving other populations when available and relevant. Salient prior findings are not

exhaustive, but reflect a sample of studies and effect size values frequently cited in the literature regarding the particular construct and rater-pair, while the findings of the present meta-analysis are more comprehensive and up to date. The comparison of these estimates reflects the difference between popular perception and a more comprehensive and cumulative summary of all available studies. In regard to other meta-analyses, no others currently available cover this particular population (i.e., specifically youth with ASD); across three different rater-pairs; with separate estimates for anxiety, depression, and broad internalizing; and cover both correlational and standardized mean difference effect size estimates in order to capture different dimensions of cross-rater agreement. Despite these differences, it is still worthwhile to assess how similar or different the inter-rater findings are across different populations and different construct variations, which will inform our understanding of how generalizable or population/construct-specific the inter-rater findings may be.

Correlation Effect Size Estimates. Previous research, as detailed in the literature review, reported youth vs. parent inter-rater correlational values for anxiety symptoms in youth with ASD ranging from -.02 to .69 (Blakeley-smith et al., 2012; Chow, 2008; Lopata et al., 2010; Magiati et al., 2014; Ooi et al., 2016); therefore, results of the present meta-analysis (average r = 0.40 [$p < 0.001$]; range = -0.02 to 0.69) remain consistent with previous literature for anxiety, given the middle value of the range constructed from salient studies in the literature is equal to .36 and this meta-analysis found a mean correlation value of 0.40. For depression, prior studies found youth vs. parent inter-rater correlations typically in the 0.29 to 0.31 range (Chow, 2008; Hurtig et al., 2009; Lopata et al., 2010), which was generally consistent with a medium effect. Though the present meta-analysis also found a medium effect size for the mean correlation, this mean correlational value and range of values reported across individual studies (r = 0.41 [$p <$

0.001]; range = -.16 to .67) were greater than anticipated based on the general estimates found to be most salient in the literature. It is not possible to know precisely why this difference occurred, however, the present meta-analysis captured broad variation in ASD studies (reflecting the full functional range within which these rater-pairs occur, not focusing only on HFASDs [e.g., Chow, 2008; Lopata et al., 2010]), was cumulative and comprehensive, and captured more recent findings. It is also possible that the prior range of values were unusually select (i.e., selection effect) and concentrated within a lower, narrower range. Finally, prior results indicated youth vs. parent inter-rater correlations concerning broad internalizing problems were generally between .25 and .56 (Hurtig et al., 2009; Jepsen et al., 2012; Kaat & Lecavalier, 2015). Present findings yielded a mean correlational value of .43 ($p < 0.001$; range = .25 to .61) for parent vs. self-reported broad internalizing, which falls within the general range of prior salient studies in the literature. That is, the present meta-analysis yielded a similar range of effect size values to the previously reviewed literature and the mean correlation for the present study was very close to the center of both range values (i.e., .25 and .56; .25 and .61). Stratis and Lecavalier's (2015) meta-analysis involving samples of youth with ASD or with ID (without ASD) yielded an overall correlation between parent vs. self-report of internalizing problems of $r = 0.42$. This value is nearly identical to the mean correlation found in the current meta-analysis ($r = 0.43$), which suggests that this correlational value is likely a good representation of the agreement between parents and youth when reporting on the internalizing symptoms of youth with ASD.

With regard to agreement between parent vs. teacher ratings, prior correlational findings—from observational studies—for anxiety symptoms ranged from .14 to .34 (Jepsen et al., 2012; Kanne et al., 2009; McDonald et al., 2016), while the present meta-analysis yielded a mean correlational value of .27 ($p < 0.001$; range = .14 to .41). For depression in youth with

ASD, previously reported correlations between parent and teacher ratings ranged from .08 to .35 (Jepsen et al., 2012; Kanne et al., 2009; McDonald et al., 2016), while the mean correlation in the present analysis was r =0 .26 ($p < 0.001$; range = 0.08 to 0.45). Finally, for broad internalizing ratings between parent and teacher, prior literature reported correlational values that ranged from 0.21 to 0.28 (Jepsen et al., 2012; McDonald et al., 2016). The present meta-analysis yielded a mean correlation of $r = 0.30$ ($p < 0.001$; range = 0.05 to 0.60) for parent vs. teacher ratings of broad internalizing. Again, these findings, across the three internalizing constructs, were generally consistent with what Stratis and Lecavalier (2015) found in their meta-analysis ($r = 0.25$) involving studies of ASD and/or ID.

It is important to note that Stratis and Lecavalier (2015) operationalized internalizing in a broad and inclusive manner in their meta-analysis--allowing either an estimate for anxiety, depression, or broader internalizing from each study to reflect the internalizing construct (i.e., whichever was the best available estimate of internalizing in each study). This allowed them to pool a larger number of studies into their overall internalizing effect size estimate, but it did not allow for them to distinguish between and report separate estimates for anxiety, depression, and broader internalizing. In taking this approach, Stratis and Lecavalier made the assumption that cross-rater correlations for anxiety, depression, and broader internalizing would be similar enough to warrant pooling together. However, they did not report any evidence of this assumption being warranted. Results of the present study, suggest that such pooling across the three constructs is reasonable, given that mean cross-rater correlations were so similar (.26 to .30) across anxiety, depression, and broader internalizing—and the similarity of those average correlations to the single mean r reported by Stratis and Lecavalier.

For all three constructs, the mean parent vs. teacher inter-rater correlations obtained from the present study were generally consistent with the middle of the effect size ranges represented among prior studies. However, based on the literature review, mean correlations were predicted to be in the medium effect size range (i.e., ≥ .30 < .50)—which is where more of the prior cited estimates tended to fall. Though statistically significant (.26 to .30 [with .296 rounded to .30 here]), the obtained average correlations from this meta-analysis fell within the small effect size range (i.e., ≥ .10 < .30)—just below the medium range minimum. In terms of precision, the 95% confidence intervals around these mean correlation values overlapped substantially with both the small and medium effect size ranges—suggesting that the true mean correlations are somewhere between a small and medium value. Additionally, highest and lowest values reported for both prior and presently reviewed coefficients reflected generally consistent ranges with respect to both anxiety and depression. However, a much wider range of correlation values was reported between parent and teacher for the internalizing construct in the present analysis, when compared to prior findings. This suggested that the present, more comprehensive, meta-analysis was likely able to capture a broader range of available estimates and, thereby, better represent more extreme values.

A specific hypothesis was not generated for the correlation between teacher report and self-report ratings of anxiety, depression, and broad internalizing, but exploratory analyses were conducted and these results were all non-significant (anxiety: $r = 0.229$ [$p = 0.090$]; depression: $r = 0.342$ [$p = 0.097$]; broad internalizing: $r = 0.316$ [$p = 0.255$]). Though not conclusive, the lack of statistical significance likely resulted from low statistical power, which is not surprising given the small number of studies available for this rater-pair (i.e., *k* ranged from 2 to 3 studies for these three statistical tests). In this regard, it is noteworthy that the obtained mean *r* values

reported for teacher vs. self-report (i.e., .229 to .342) were generally consistent with the mean values observed for the parent vs. teacher rater-pair (i.e., mean *r* ranging from .256 to .296)—though the teacher vs. self-report estimates varied slightly more, the mid-range value of approximately .28 is common to both sets of average rater-pair correlations. Further, the average effect size estimate of the correlation between teacher vs. self-reported broad internalizing resulting from this meta-analysis ($r = 0.316$) was fairly consistent with the effect size estimate reported by Stratis and Lecavalier (2015; $r = 0.25$), and the mid-range $r = 0.28$ value across the three internalizing constructs in the present study was even closer.

Mean Difference Effect Size Estimates. Prior observational studies have found that, when compared to the self-reports of youth with ASD, parents tend to report higher levels of anxiety, depression, and broad internalizing problems in youth with ASD (Barnhill et al., 2000; Bitsika & Sharpley, 2015; Kaat, 2014; Lopata et al., 2010; Taylor et al., 2018). Similarly, research by Barnhill et al. (2000) indicated that teachers also report higher levels of youth anxiety (teacher $M = 60.10$; youth $M = 47.19$) and depression (teacher $M = 62.00$; youth $M = 50.56$) compared to what youth with ASD report themselves. Further, prior literature indicated that, when rating youth with ASD, parents and teachers tend to report similar mean levels of anxiety, depression, and broad internalizing (Barnhill et al., 2000; McDonald et al., 2016). In contrast to these ASD-related findings, Huang (2017) conducted a meta-analysis regarding broad behavior issues and involving typically developing and clinical samples of youth. This study revealed that, when measuring broad internalizing problems, the overall effect size estimate between parents and youth was $g = -0.21$ (with youth reporting more internalizing problems than parents), the effect size estimate between teachers and youth was $g = -0.76$ (with youth reporting

more symptoms than teachers), and the effect size estimate between parents and teachers was g = 0.52 (with parents reporting more youth internalizing problems than teachers).

Consistent with the above prior research findings from observation studies involving samples of youth with ASD, the present meta-analysis found that the significant overall mean differences between parent and self-ratings for anxiety (g = 0.220) and depression (g = 0.788) reflected parent report yielding a higher mean (i.e., more perceived symptoms of anxiety and depression) than youth self-report. Conversely, the overall mean difference between parent and youth self-report of broad internalizing was near zero (g = 0.090) and not significant. Thus, these parent vs. youth self-report results aligned with prior ASD research for mean differences in levels of reported anxiety and depression, but not for broad internalizing issues. However, the non-significant, near zero, overall result for broad internalizing in the present meta-analysis was based on only four available studies. This may be too few studies to draw firm conclusions. It is possible that the mean effect size estimate was not stable and perhaps additional studies would alter the estimate, but such a hypothesis cannot be examined without additional effect size estimates from future studies. Of note, results of the present meta-analysis were not consistent with the findings of Huang's (2017) meta-analysis, which examined inter-rater differences in a broad range of typically developing and clinical youth samples using the CBCL. Specifically, Huang (2017) found that youth tended to rate themselves higher on internalizing symptoms than their parents rated them, while results of the present meta-analysis indicated the opposite rater pattern for both anxiety and depression, and yielded a near-zero mean difference for broad internalizing. This difference in findings is likely due to the current meta-analysis involving exclusively ASD samples, where youth with ASD are likely to have more difficulty reporting on

their own internalizing emotions (Baron-Cohen et a.,1985; Bird & Cook, 2013; Kiep & Spek, 2016).

Application of Duval and Tweedie's Trim and Fill method (Borenstein et al., 2014) to the parent vs. self-report rating standardized mean difference effect size estimates yielded evidence of potential publication bias in the available mean difference effect size estimates for anxiety and depression. The analysis suggested the possibility that studies with smaller samples were more likely to be published when parent ratings were significantly higher than youth self-report ratings. When the overall standardized mean difference effect size estimates were adjusted for anxiety and depression, the anxiety estimate moved from a small significant effect size (0.21963) favoring parent ratings to a non-significant near-zero estimate (0.05815) and the depression estimate moved from a medium significant effect size (0.78760) favoring parentings to a small non-significant effect size (0.24430). In the case of anxiety, the bias adjustment does not result in a substantive change in interpretation, as the original unadjusted small effect size estimate was not particularly meaningful from a clinical perspective. However, the difference between the original and bias-adjusted estimate for depression is a more substantial concern. The original difference was just under the minimum for a large effect--with the 95% confidence interval overlapping with the both the medium and large effect size ranges. The adjustment results in a drop in the overall estimate of more than half of a standard deviation. Whether an adjustment this extreme is warranted is not clear, but the estimated increase in variability of the effect size estimates introduced by the fill part of the trim and fill method appeared too extreme to be considered reasonable. As a result, though the original overall standardized mean difference for depression is likely inflated, the extent of the bias adjustment should be interpreted with caution. This finding suggests that a closer examination of this particular part of the literature (i.e.,

regarding mean differences in parent vs. youth self-ratings of depression, both within ASD and in other samples) is warranted going forward.

Differing from prior literature, the mean differences between teacher and self-rated mean values of anxiety (g = 0.295 [p = 0.417], k = 6), depression (g = 0.670 [p = 0.097], k = 4), and broad internalizing (g = -0.033 [p = 0.930], k = 2) did not significantly differ. However, within the teacher vs. self-rated anxiety and depression constructs, there was unusually large variability in the effect sizes reported by the studies included in this analysis (g range = -0.035 to 3.314 for anxiety, and g range = -0.334 to 1.683 for depression).. The combination of small numbers of studies involved and unusually large variability in effect size estimates, suggests that the overall effect size values were likely unstable estimates of the true effect sizes and that statistical power was likely an issue for significance testing. In addition, the variability in effect size estimates may be suggestive of possible moderators, which could not be presently assessed due to the lack of sufficient studies for this type of analysis. Regarding mean differences in the teacher vs. self-reported broad internalizing construct, there were only two studies represented in the analysis, which makes it problematic to perform a statistical test or draw broad conclusions. Further, during the initial literature search, only one observational study (Barnhill et al., 2000) was found regarding this rater-pair out of all three constructs. As a result, hypothesis 5b was generated based on only this study and additional theoretical evidence. With a more comprehensive search strategy, more observational studies became available to be included in this analysis and it seems that the Barnhill et al. (2000) study was less representative of the broader literature. Overall, more studies are needed for this rater pair before it will be possible to draw firm conclusions about the mean differences between teacher vs. self-reported anxiety, depression, and broad internalizing in youth with ASD.

Within the parent and teacher rater-pair, findings from this meta-analysis were consistent with previous ASD literature. That is, mean teacher ratings did not significantly differ from mean parent ratings of youth for depression (g = 0.176 [p = 0.349], k = 9) and, although mean differences did significantly differ for anxiety (g = 0.156 [p = 0.002], k = 11) and broad internalizing (g = 0.153 [p = 0.041], k = 14), the effect size estimates remained negligible In contrast, the present study results were inconsistent with results of Huang's (2017) meta-analysis which found a moderate standardized mean difference between parent and teacher internalizing ratings—with parents giving higher ratings. However, again Huang's meta-analysis pertained to broader typically developing and clinical samples--and was not about ratings in an ASD context. Thus, mean difference results between raters may be different in the ASD context relative to broader typically developing and clinical samples—and this may even be the case when self-reports are not involved. Finally, it is important to note that there was moderate-large variability in the reported effect sizes for studies included in these analyses (i.e., anxiety g range = -0.005 to 1.371; depression g range = -0.066 to 0.964; internalizing g range = -0.028 to 0.656). This could be suggestive of possible unknown moderators and should be studied more in future studies.

Impact of Cognitive Ability

Correlation Effect Size Estimates. In general, prior literature found higher agreement between youth self-report and parent report of internalizing problems when broad samples of youth (i.e., not specifically ASD samples) had higher verbal or cognitive abilities (Durbin, 2010; Vasa et al., 2016). However, the results of this meta-analysis found that the correlations between youth self-report and parent report of anxiety, depression, and broad internalizing were not significantly moderated by the general cognitive ability of the youth (mean youth FSIQ value). Yet, the p value of .0502 for FSIQ as a moderator for youth self-report vs. parent ratings of

anxiety should be noted. When the scatter plot for this analysis was examined (see Figure 18 in Appendix G), an unusual but potentially meaningful pattern was observed. The coefficients appeared to form two distinct groups (Group 1 with mean FSIQ between 87.78 and 94.98 and Group 2 with mean FSIQ between 101.66 and 110.14). Group 1 appeared to show a positive trend indicative of the correlation increasing as the mean FSIQ increased, while Group 2 appeared to show a less steep negative trend where the correlation appeared to decrease to near 0 or even slightly negative as the mean FSIQ increased. This pattern was suggestive of a potential curvilinear relationship between inter-rater agreement for youth self-report vs. parent report, in the context of youth cognitive ability, when rating anxiety. When the two groups were analyzed separately, the slopes with different directions did appear. Neither achieved statistical significance, but the number of studies in each group was considerably smaller (Group 1 $k = 6$ and Group 2 $k = 7$) than that for the overall, combined moderator analysis. (Note that despite the small sample size, the test of the moderator for Group 2 achieved a p value of .0599.) Ultimately, the lack of statistical significance left any final conclusion in doubt regarding youth self-report vs. parent ratings of anxiety being moderated by FSIQ. However, future studies should assess for a potential curvilinear or negative overall relationship between IQ and inter-rater correlations for anxiety in the youth self-report vs. parent rating context. Additionally, a potential outlier in the depression scatter plot was removed and the moderator analysis rerun. However, this quite clearly did not alter the non-significant conclusion. Finally, the moderator analysis for broad internalizing involved only four studies. There was no trend apparent in the scatter plot (see Figure 23 in Appendix G), but any true mild trend could be overwhelmed by sampling error with so few studies. In the absence of other evidence, mean FSIQ does not

appear to moderate the correlation between youth self-report and parent ratings of depression or broad internalizing.

Analyses to investigate youth self-report vs. teacher ratings could not be completed due to a lack of studies in this category that reported youth cognitive ability (k = 1 for anxiety, k = 2 for depression, and k = 1 for broad internalizing) for their samples. Clearly, more studies that report IQ information are needed to better understand if FSIQ acts as a moderator in the context of this rater-pair.

It is interesting to note that available studies reporting mean FSIQ for youth with ASD in the self-report-related inter-rater context, reported mean FSIQs ranging only from the low average to high average ranges. Despite the frequency of comorbid ID within the context of ASD, available studies did not appear to focus specifically on the ID range of functioning when self-report rating was required (though some did include a minority of cases within the ID range in the context of an otherwise higher-functioning sample). This suggests that ASD researchers appeared to typically avoid use of self-report ratings of internalizing states in the ID context. This may have been due to the belief that cases in the ID range with ASD would be unlikely to provide useful self-ratings of internalizing issues. However, the lack of such cases being represented among the studies may have reduced the range of talent for tests of FSIQ as a potential moderator, as the expectation that ratings from such cases would typically provide low or negligible agreement with other raters is consistent with the prediction that FSIQ would be positively related to inter-rater agreement when self-report ratings are involved. Thus, the anticipated relationship may have been undermined by excluding samples made up of cases representing the lower end of the FSIQ distribution. Another likely issue for those within the ID range is the greater likelihood of insufficient reading skills to complete the self-rating protocol

independently (Ratz, 2013)—though items could be read to those with insufficient reading skills. Yet, though not a part of standardized procedures for most self-report rating scales, this is likely often done in practice, when needed—and some studies did allow for this (e.g., Adams et al., 2013; Bellini, 2004; Blakely-Smith, et al., 2012; Chow, 2008). However, the potential need for this accommodation may have discouraged some researchers from including such cases.

As an exploratory analysis, FSIQ moderation for mean correlations between parent and teacher report of youth depression and broad internalizing was examined. (There was no specific prediction to be made regarding if or how cognitive functioning of the child could impact agreement between ratings by adults who regularly interact with the child.) Results indicated that the mean correlations between parent and teacher report were not significantly moderated by youth cognitive ability for depression or broad internalizing. Given the small number of studies, the fact that the initial heterogeneity tests indicated that there was not any substantive true variation in effect size estimates for a moderator to account for, and the lack of a theoretical justification for why FSIQ might moderate in this context, these results are not surprising. For anxiety, there were not enough studies available to complete this moderator analysis in the parent vs. teacher rating context ($k = 3$). Clearly, additional future studies of parent vs. teacher report of internalizing issues that report IQ information would be helpful in terms of improving precision and power—and allowing for a stronger moderation test. Yet, the lack of theoretical justification makes this less of an urgent need relative to other questions.

Mean Difference Effect Size Estimates. Based on previous literature, and as mentioned above, research supports that general cognitive ability is associated with better agreement between parent and youth self-reports of anxiety, depression, and broad internalizing (Blakeley-Smith et al., 2011; Durbin, 2010; Kaat & Lecavalier, 2015; Ooi et al., 2016; Vasa et al., 2016).

By extension, as in the correlation context, it is reasonable to make the same assumption with the relationship between teacher and youth self-reporters. Further, no prior studies were identified that investigated youth cognitive ability as a moderator between parent and teacher ratings of these internalizing problems.

Inconsistent with what was predicted based on prior research findings involving samples of youth with ASD, this meta-analysis found that the mean differences between youth self-report ratings and parent ratings of anxiety and depression were not significantly moderated by youth cognitive ability (mean FSIQ value). The scatter plots (Figures 32 and 33 in Appendix G) for these analyses were examined to determine if anything stood out as a reason why the findings were not as expected. Both scatter plots show a slight, positive trend indicating that the higher the youth FSIQ, the bigger the mean differences. One possible explanation for this is that parents may be more involved with the process of youth completing the measure of self-report when youth have lower cognitive abilities. Thus, there could be less independence within the rater-pair, leading to fewer mean differences. Further, youth cognitive ability was not found to significantly moderate the mean differences between teacher and self-report of anxiety, although there were only four studies involved in this analysis, which may be why it was not significant. This moderator analysis was unable to be completed for the mean differences between parent and youth self-report of broad internalizing or teacher and youth self-report of depression or broad internalizing because too few studies reported information on youth cognitive ability for their samples. Again, as stated in the correlation context, studies need to typically report information on youth IQ in order to evaluate cognitive ability as a potential moderator variable for inter-rater agreement.

As an exploratory analysis, cognitive ability of youth as a moderator of the mean differences between parent and teacher report ratings was completed; however, cognitive ability was not found to significantly moderate mean differences between parent and teacher ratings of anxiety, depression, or broad internalizing. This result can be expected, as no reasonable hypothesis was able to be generated regarding an effect of youth cognitive ability on the relationship between parent and teacher report of youth anxiety, depression, and broad internalizing.

Impact of Age

Correlation Effect Size Estimates. Prior research findings supported higher agreement between youth self-report and parent report of anxiety, depression, and broad internalizing in older adolescents compared to younger children (Achenbach et al., 1987 [broad sample of youth]; Ebesutani et al., 2011 [typically developing sample of youth]; Stratis & Lecavalier, 2015 [ASD or ID without ASD sample of youth]). It was also hypothesized that, due to the possibly different presentation of internalizing problems in young children (e.g., reflected in temper tantrums or misbehavior; Frick et al., 1994), it would be more difficult for parents and teachers to produce convergent ratings of internalizing problems when rating younger children than when rating older adolescents.

However, results of the present meta-analysis, involving all available and relevant studies, found that mean correlation values between youth self-report and parent report ratings were not significantly moderated by mean age for anxiety, depression, or broad internalizing. Thus, results for all three constructs apparently failed to support this hypothesis. Given the unexpected nature of this result, the scatter plots for all three constructs were closely examined (see Figures 26 and 27 in Appendix G). This revealed an outlier in the depression scatterplot.

When this outlier was removed, the significant and substantive positive relationship between mean age and the inter-rater depression correlations was apparent. Thus, the youth self-report vs. parent ratings correlation value for depression was positively related to, and moderated by, the mean age of the study sample, which was consistent with prior findings in other meta-analyses (Achenbach et al., 1987 [broad sample of youth]; Stratis & Lecavalier, 2015 [sample of youth with ASD or ID without ASD]). It is important to note that in these prior meta-analyses, constructs were combined to reflect general estimates of internalizing problems, while in the present meta-analysis, anxiety, depression, and broad internalizing constructs were analyzed separately—and age of youth was a significant moderator for only the depression construct. Given this result, it is possible that the overall trend observed by Achenbach et al. (1987) and Stratis and Lecavalier (2015) could have been largely due to depression effect size estimates, but this would not have been clear to them due to pooling effect size estimates from anxiety, depression, and possibly broader internalizing together.

Findings for mean age as a moderator were generally non-substantive for the youth self-report vs. teacher report rater-pair and the parent vs. teacher rater-pair. In the case of youth self-report vs. teacher report ratings, there were not sufficient relevant studies to conduct the moderator analysis for any of the three constructs (*k* ranged from 2 to 3 studies for each construct). For the parent vs. teacher rater-pair, the number of relevant studies available ranged from six (anxiety) to nine (broad internalizing) studies. In this rater-pair context, the moderator analyses for all three constructs were clearly non-significant with non-substantive slopes and R^2_{Analog} results. Thus, mean age does not appear to moderate the correlation between raters in the parent vs. teacher rater-pair context, and lack of sufficient data prevented any conclusions regarding mean age as a moderator in the youth self-report vs. teacher report inter-rater context.

Mean Difference Effect Size Estimates. Huang (2017), using a broad range of typical and clinical samples, found that mean differences between parent report and youth self-report were significant for younger children (g = -1.07), but not older adolescents (g = -0.15) when rating youth broad internalizing problems. Additionally, as stated when discussing research question 3, there is evidence to support older adolescents having more developed skills that could increase their accuracy of self-report (Hill, 2004; Kiep & Spek, 2016; Spek, et al., 2009) and, potentially, lead to smaller mean differences between their own self-report compared to parent and teacher reports. Again, using the same logic as in the correlation context (i.e., as with research question 3), variation in the outward expression of internalizing problems in young children (Frick et al., 1994) could make it more difficult for third-party raters (e.g., parents and teachers) to judge and report similar levels of internalizing problems when rating young children compared to when rating older adolescents.

Unlike what was expected based on prior research findings, this meta-analysis found that the mean differences between youth self-report and parent report were not significantly moderated by youth age (mean age in years) for anxiety or depression. However, visual examination of the scatter plots (see Figures 35 and 36 in Appendix G) did reveal a negative trendline suggesting smaller cross-rater mean differences as age increased. It is possible that with additional studies, if the trend holds and statistical power increases, these analyses would become significant. Similarly, mean differences between youth self-report and teacher report for anxiety were also not significantly moderated by youth age. However, it is important to highlight that this analysis included only five studies, which works against making generalizable conclusions regarding moderation until more studies are available to improve estimate stability and power. Finally, the mean differences between parent and teacher ratings for anxiety,

depression, and broad internalizing were not significantly moderated by youth age. Given too few available studies, this moderator analysis could not be run at all for youth self-report vs. parent report of broad internalizing or youth self-report vs. teacher report of depression and broad internalizing. Overall, it was concluded that age of the youth was not significant moderator of the mean differences between the raters. However, all age-related moderator analyses in the mean difference context were plagued by too few studies available for robust estimates and low statistical power. Thus, the potential influence of youth age should be examined further in future individual studies involving larger sample sizes and, in meta-analyses, as more studies become available.

Impact of Method of Self-Report Administration

Correlation Effect Size Estimates. Though not explicitly acknowledged in the prior research literature, a detailed examination of studies for the literature review suggested a possibly cross-rater trend in relation to the conditions under which the youth self-report ratings were obtained. A number of studies in this prior research reported relatively high levels of agreement between parent and youth report ratings of internalizing problems when the self-report rating scales were completed in the home setting (Bitsika et al., 2016, Farrugia & Hudson, 2006; Jepsen et al., 2012; Magiati et al., 2014), while generally poorer levels of agreement were reported when the self-report rating scales were completed in a clinic setting (e.g., Chow, 2008; Lopata et al., 2010; Taylor et al., 2018).

Initial coding of the studies for the meta-analysis placed them in the following general categories that emerged during the review: (a) assessment read to the child in the clinic, (b) assessment completed in the clinic, (c) assessment read to the child at home, and (d) assessment completed at home. Ideally, it would have been possible to use a more precise set of coding

categories for the method (i.e., setting and administration) used to complete self-report ratings. But these four categories were as precise as possible, given the information provided in the studies.

Though all four categories contained multiple articles overall, no single moderator analysis contained articles that reflected all four categories—and in a number of cases, one of the available categories contained too few articles to be meaningfully used as a category for comparison purposes. Ultimately, for the sake of maximizing statistical power, while still allowing for the comparison of meaningful categories, the four coding categories were collapsed into two coding categories (i.e., the two categories involving home completion were collapsed into "assessment completed at home" and the two categories involving clinic completion were collapsed into "assessment completed in clinic").

Examination of this two-category method of self-report rating variable as a potential moderator of the correlation between youth self-report vs. parent report ratings for anxiety yielded a strong moderation effect. For these two categories, the "assessment completed in clinic" category yielded a mean $r = 0.319$ (a medium effect size) and the "assessment completed at home" category produced a mean $r = 0.559$ (a large effect size).

This result deserves some further discussion concerning its possible meanings. Though one may be tempted to interpret the higher correlation in the home context as a positive finding (e.g., possibly reflecting a youth's greater comfort in the home setting), this is not necessarily the case. For example, it could mean that parents are more likely to assist the children in completing the ratings at home, rendering the completion of ratings less independent for the rater-pair. This would serve to spuriously increase the correlation as a result of non-independence in rating completion. It may also indicate that children could be responding, when supervised by parents,

in a manner consistent with what they believe their parents want and, therefore, they are less likely to answer honestly knowing that their parent may see their responses. Although these are not conclusive interpretations, and this is not an exhaustive list of possibilities, it is noteworthy that these are all potentially testable hypotheses for future research. However, the tentative interpretation should be the suspicion that completion of self-report ratings in the home setting is likely to lead to a spuriously higher correlation resulting from partial lack of independence between the raters during ratings completion. If this is true, then the clinician overseeing the self-report ratings is the likely solution. If the child is also less comfortable in the clinical setting, then it is possible that a clinician could administer the self-report ratings at the child's home (though feasibility may be an issue in this case). Potential solutions will depend on a better understanding of the nature of the problem, which was partially revealed through this moderator analysis. This problem involving method of self-report ratings completion should be a concern for both researchers and clinicians. At the very least this finding should be taken into account, but once the underlying reasons for it are better understood, researchers and clinicians may need to take explicit steps to minimize the discrepancy in a manner consistent with both use of independent sources of information in assessment and the validity of ratings.

The findings for this follow-up analysis in the context of depression were less revealing. That is, results were not significant. This was not surprising due to the initial test of heterogeneity indicating that there was no significant variation in true effect sizes available to be accounted for by moderators. However, the relatively small number of studies ($k = 7$) may have been a factor in this result (i.e., restricting variability and limiting power). Consistent with this possibility,, the mean correlation values did follow the established pattern of stronger agreement when the assessment was completed at home ($r = 0.361$ [$k = 5$] for "assessment completed in

clinic" and $r = 0.501$ [$k = 2$] for "assessment completed at home"). Unfortunately, in the case of broad internalizing, the small number of studies ($k = 4$) made a moderator analysis untenable.

Given the strong finding for anxiety and the limited number of studies available for depression and broader internalizing, all three of these constructs should be assessed in future studies involving a larger number of studies, greater statistical power, and, ideally, more refined self-report method and context categories (resulting from more precise and detailed reporting by study authors).

Mean Difference Effect Size Estimates. As previously described, careful examination of prior observational studies involving samples of youth with ASD indicated a trend related to the method of self-report administration and level of agreement. Specifically, more convergent parent and youth self-report mean level ratings of anxiety, depression, and broad internalizing were observed when the assessment was completed in the home setting (Bitsika et al., 2016; Farrugia & Hudson, 2006; Jepsen et al., 2012; and Magiati et al., 2014). Conversely, more divergent mean differences were found when youth were read the items of the rating scale in a clinic setting (Chow, 2008; Lopata et al., 2010; Taylor et al., 2018).

However, in this meta-analysis, the mean differences between youth and parent ratings of anxiety and the mean differences between youth and parent ratings of depression were not significantly moderated by the method of self-report, and there were not enough studies ($k = 3$) within the broad internalizing construct to conduct this moderator analysis at all. However, when the four self-report method categories were collapsed into two categories ("assessment completed in clinic" and "assessment completed at home"), the category "assessment completed at home" resulted in smaller mean differences (i.e., parent vs. self-reported anxiety: $g = 0.131$; parent vs. self-reported depression: $g = 0.412$) than the category "assessment completed in

clinic" (i.e., parent vs. self-reported anxiety: g = 0.384; parent vs. self-reported depression: g = 0.984). Based on these values, there is a fairly large difference between the mean effect size values within these categories. As previously discussed in the correlation context (i.e., regarding research question 4)—smaller mean differences when the assessments were completed in the home setting may be a result of direct or indirect parent influence on child self-report (i.e., non-independence of the raters), leading to greater, but spurious, agreement between the parent/child pairs and should not necessarily be considered a positive finding. The meaning of this finding clearly requires further exploration in future research.

Correlation and Mean Difference Results by Rater-Pair and Construct

The results across inter-rater correlations and standardized mean differences will be summarized together and discussed by rater-pair and construct. The intent of this section is to clarify all major results for each construct within a given rater-pair, as discussing these results separately across research questions can conceal how they connect or otherwise relate to each other. In addition to having the potential to make overall patterns clearer, this approach can also provide insight regarding where the needs are for additional individual studies going forward that could improve precision and make more robust analyses possible. These results are summarized and organized by rater-pair and construct in Table 27.

Parent vs. Self-Report: Anxiety. Within the parent vs. self-report of anxiety in youth with ASD, this meta-analysis found that the mean correlation (r = 0.399) was significant and medium in effect while mean differences were also significant, but small in effect (g = 0.220]) with parent ratings yielding higher means than self-ratings. This means that while agreement was medium in effect based on the mean correlation effect size estimate, there were still significant mean differences between parent and youth self-report of anxiety. However, it is

important to, again, highlight the presence of potential bias regarding the mean difference effect size estimate within this rater-pair and construct. Because the bias-adjusted effect size estimate moved closer to 0 (i.e., $g = 0.058$, non-significant), it is possible that the true mean difference between parent and youth self-report ratings of anxiety is smaller than the unadjusted effect size estimate would suggest—and perhaps negligible. Further, within this rater-pair and construct, neither mean *r* nor mean *g* were significantly moderated by cognitive ability of the youth. However, the borderline *p* value ($p = 0.0502$) and potentially substantive R^2_{Analog} of .30 for cognitive ability as a moderator, in the inter-rater correlation context, leads to the suspicion that statistical power may have played a role in this moderator failing to achieve statistical significance. Even more intriguing is the observation that the scatter plot for FSIQ (in the context of the inter-rater correlation) suggested a possible curvilinear relationship—wherein the slope is positive from lower to average IQ and negative or flat in the average to high average range. This is something that should certainly be examined in other studies going forward—under conditions of greater power and use of curve-fitting strategies. Similarly, age of youth was not found to significantly moderate mean correlations or mean differences. The mean correlation between parent report and self-report of anxiety was found to be moderated by the method of self-report administration with the strongest correlation occurring in the "assessment completed at home" category (assessment completed at home [$r = 0.559$], assessment completed in clinic [$r = 0.319$]). However, within the analysis of standardized mean differences, method of self-report administration was not a significant moderator variable. Yet, it is important to note that the above pattern for correlation was mirrored in this mean difference moderator analysis (though not statistically significant) with smaller mean difference effect sizes occurring within the "assessment completed at home" category (i.e., assessment completed at home [$g = 0.131$],

assessment completed in clinic [g = 0.384]). (The failure to find a significant difference between the negligible g for "assessment completed at home" and the small, though potentially substantive, g for "assessment completed in the clinic" may have resulted from low statistical power for detecting a difference in the lower portion of the correlation matric. Thus, the potential for this moderator, in this context, has not been completely ruled out by the non-significant result.)

Parent vs. Self-Report: Depression. Regarding the parent vs. self-report ratings of depression, the observed mean correlation (r = 0.412) and the observed standardized mean difference (g = 0.788) were both significant and reflected a medium effect in their respective, r and g, metrics. The observed g result, though of medium magnitude, was near the minimum large effect size standard of 0.80. Thus, this finding reflected a fairly substantial difference--wherein, on average, parents tended to rate depression in the youths as much higher than the youth's rated themselves. However, given significant evidence of possible publication bias, the bias-adjusted g estimated for depression was considerably smaller than the observed, unadjusted value and non-significant (g = 0.244. Thus, the true effect size g for depression in the parent vs. youth self-report rating context is likely much closer to 0 than the observed value based on the available studies. If this is true, the bias-adjusted value suggests closer convergence of mean parent and mean youth self-report ratings. Though the g of 0.244 reflects a small effect size, its bias-adjusted 95% confidence interval suggests the possibility that the true g, like the bias-adjusted value for anxiety, may be negligible. Further research is needed to clarify this potential bias issue in the mean difference ratings for parent vs. youth self-report ratings of depression. As indicated previously, the bias-adjusted increase in the estimated variability of the effect size g

distribution appeared unreasonably large in the depression context for this rater pair—leading to questions about possible correction excess.

Within this rater-pair and construct, neither mean correlations nor mean differences were significantly moderated by cognitive ability of youth or age of the youth. However, upon removal of an outlier, age of the youth was found to significantly moderate the inter-rater correlation—clearly indicating that as age increased, the correlation between parent and youth depression self-ratings increased. Although the mean correlation and standardized mean difference between parent and self-ratings of depression were not significantly moderated by the method of self-report administration, their obtained values for the two conditions did follow the predicted pattern—leaving open the possibility that the non-significant results may have been due more to attenuated statistical power than true lack of moderation. When the self-report ratings were completed at home, the mean *r* was 0.501 and the mean *g* was 0.412, but when the self-report ratings were completed at the clinic, the mean *r* was .0361 and the mean *g* was 0.984. Though the differences did not reach statistical significance, the observed results for both *r* and *g* were consistent with stronger agreement when self-ratings were completed at home (higher *r* [large] and lower *g* [small]) than when completed in a clinical setting (lower *r* [medium] and higher *g* [large]).

As indicated previously, it is not clear exactly why greater rating convergence may tend to occur at home compared to the clinic. However, it is suspected that possible assistance from or influence of the parent in the home setting renders completion of the self-ratings non-independent of the parent ratings. If true, this would spuriously increase the agreement between parent and youth ratings. In contrast, it is possible that the youth are more comfortable at home than in the clinical setting—and that the relative discomfort in the clinical setting somehow

adversely impacted the ratings. Other explanations are also possible (e.g., whether differences are due to the influence of the actual setting, due to a person in that setting, due the actions of a person in that setting; whether or not the items are read by the youth or read to the youth by someone else, etc.). In general, these different possible explanations are potentially testable if such variables were to be appropriately assessed, recorded, and reported in individual studies. Alternatively, individual studies could be set up that manipulate these different conditions to assess their potential impact and systematically rule out rival explanations.

Parent vs. Self-Report: Broad internalizing. Within the construct of parent vs. self-ratings of broad internalizing, the mean correlation was moderate and significant (r = 0.430), while the overall standardized mean difference was negligible and non-significant (g = 0.090). In general, the overall medium r and negligible g for internalizing match up closely with the mean r and bias-adjusted g results for both anxiety and depression. However, viable moderator tests for FSIQ of the youth, mean age of the youth, and method of self-report completion were not possible for broad internalizing due to the small number of available studies.

Parent vs. Teacher Report. Results for anxiety (r = 0.273, g = 0.156), depression (r = 0.256, g = 0.176), and broad internalizing (r = 0.296, g = 0.153) in the parent vs. teacher rating context followed each other closely. For all three constructs, the observed mean r was significant, fell within the small effect size range, and the 95% confidence interval overlapped with the small and medium effect size range. In the case of the three observed overall g values, all were in the negligible range--though the p values for anxiety and broad internalizing did achieve statistical significance. Neither FSIQ nor mean age were significant moderators for any of the three internalizing constructs in the parent vs. teacher rating context—and this was the case for both mean r and g values. However, it should be noted that the number of studies was

typically insufficient for adequate moderator tests (*k* ranged from 3 to 12 studies, *Mdn* and *Mo* = 7 studies) in the parent vs. teacher rating context. A larger number of individual studies involving parent vs. teacher ratings, with clearly reported moderator options, would be helpful for future meta-analyses.

Teacher vs. Self-Report. For teacher vs. youth self-report ratings, mean observed correlations were similar and non-significant for all three constructs (i.e., anxiety $r = 0.229$, depression $r = 0.342$, and broad internalizing $r = 0.316$, ranging from small to medium effect sizes), while average observed *g* results were more variable (i.e., anxiety $g = 0.295$, depression $g = 0.670$, and broad internalizing $g = -0.033$), but still all non-significant. The lack of statistical significance is not surprising given the number of studies available for each statistical test ranged from $k = 2$ to 6 studies (*Mdn* = 2.5 studies). The same issue of insufficient studies available pertained to all moderator analyses in the teacher vs. youth self-report ratings context—with the number of studies available per test ranging from $k = 1$ to 5 studies (*Mdn* = 2 studies). Thus, only moderator analyses within the anxiety construct could be run for this rater-pair—but with still too few studies for adequate power or robust results. The clear issue going forward for this rater-pair is the need for more studies that examine and report inter-rater agreement results (*r* and *g*) and clearly report useful moderator options.

Further Considerations Regarding Moderator Variables

Continuous vs. Categorical Moderators. For this meta-analysis, cognitive ability of the youth (i.e., mean FSIQ score) and age of the youth (i.e., mean age) were both evaluated as continuous moderator variables. A prior meta-analysis (Stratis & Lecavalier, 2015) that looked at informant agreement using ASD and ID samples considered youth cognitive ability and youth age as both categorical and continuous moderator variables. Regarding cognitive ability of the

youth, Stratis and Lecavalier (2015) utilized a FSIQ cutoff score of 70 (i.e., > 70 = non-ID and < 70 = ID range) to delineate the two categorial groups. For age of the youth, these authors created three categories (i.e., preschool, school-aged, and adolescent). For both of these moderator variables, the authors reported that the studies included in their meta-analysis often reported large ranges of FSIQ and age in their samples or did not report a range at all. Therefore, in order to gather enough studies to analyze these variables as categorial moderators, Stratis and Lecavalier (2015) had to collapse across all rater types to populate the categories with sufficient studies. That is, all rater-pair groups (e.g., parent vs. self, parent vs. teacher, etc.) were pooled together in these categorical moderator analyses. For the purposes of the current meta-analysis, running these variables as categorial and continuous moderators was considered. However, the same issues described by Stratis and Lecavalier (2015) were encountered (i.e., ranges too broad to fit into one category or no ranges reported), which resulted in too few studies per rater-pair type. Because the purpose of this meta-analysis was to examine different types of rater-pairs separately, while remaining open to potential similarities and differences between them—it was not useful to follow the Stratis and Lecavalier approach of collapsing across rater-pairs and running one categorical moderator analysis per variable. Thus, cognitive ability of youth and age of youth were analyzed only as continuous variables in the present meta-analysis.

Method of Self-Report Administration Issues. As a follow-up analysis regarding the method of self-report administration, the original four categories "assessment completed in clinic," "assessment read to child in clinic," "assessment read to child at home," and "assessment completed at home" were collapsed into two categories: "assessment completed in clinic" and "assessment completed at home." This was done to gain more statistical power for the comparison and because very few studies specifically reported on whether the assessment was

read to the child. Collapsing these four categories into two did not change the statistical conclusion for any of the analyses completed; however, it did help to clearly identify the pattern that when assessments were completed in the home setting, parent/child agreement was better (in terms of higher correlational value and smaller mean differences). This moderator variable has potential to significantly impact inter-rater agreement. Therefore, it should be carefully considered in future research by specifically reporting on how and where the self-report rating scales were completed.

Type of Correlation Coefficient. Correlation type was also evaluated as a potential moderator variable, which also informed the decision to combine all correlation type values for the major analyses. As previously described, Pearson's *r*, Spearman's *rho*, and ICC were all treated as Pearson's *r* for purposes of the major analyses. Although this is not the most ideal way to treat differing correlation types (as there are some differences in the measurement dimensions covered), this approach was done more out of necessity to increase the number of studies that could be included in a single analysis and thereby increase statistical power. Because of this combining method, correlation type was analyzed as a potential moderator variable to determine if there were substantial differences in correlation effect size estimates related to correlation type. Analyses revealed that treating all correlation estimates as if they were Pearson's *r* did not significantly impact the average correlation estimates. However, to improve options for analyzing the different correlation coefficients separately moving forward, it would be helpful for studies to report all three correlation coefficients when publishing studies regarding inter-rater agreement.

Other Potential Moderators. Other potentially relevant moderator variables (i.e., score type, parent SES, ethnicity/race, gender, social desirability, parental depression, and parental

stress) were considered for examination within the present meta-analysis. However, for various reasons, it was not possible to analyze them effectively as potential moderators. First, information concerning variables such as parent SES, ethnicity/race, social desirability, parental depression, and parental stress was not consistently reported in sufficient studies to reasonably allow for meaningful moderator analyses. Second, available studies rarely separated results by gender or even represented females with ASD well. Therefore, a moderator analysis for gender could not be completed. Finally, most studies reported only standard scores with very few reporting raw scores (e.g., Bitsika et al., 2016; Keith et al., 2019; Rump, 2010; Sharpley et al., 2015), so there was insufficient variability for investigating this score distinction as a potential moderator. If would be helpful for studies to report both types of scores, when possible, moving forward.

Effect Size Estimates in the ASD Population Compared to Other Populations of Youth

As previously described, some prior meta-analyses investigating inter-rater agreement for internalizing problems in youth did not utilize exclusively ASD samples. It is helpful to compare the results of the present meta-analysis to these others in order to better understand how generalizable or population-specific such result may be.

To begin, Stratis and Lecavalier (2015) utilized a mixed sample of youth with ASD and youth with ID (without ASD) and investigated inter-rater agreement (correlation effect size) of internalizing problems as a general construct. Next, Huang's (2017) meta-analysis included studies with various typically developing and clinical youth samples and investigated inter-rater agreement of broad internalizing problems using both correlation and mean difference effect size estimates. For the most relevant comparisons, correlation and mean difference effect sizes for broad internalizing problems derived from this meta-analysis will be used.

Within the parent vs. self rater-pair, Stratis and Lecavalier (2015) found a mean value of $r = 0.42$, Huang (2017) reported a value of $r = 0.33$, and the current study yielded a value of $r = 0.43$. These internalizing correlation effect size estimates were moderate across all three studies. For the parent vs. teacher rater-pair, this meta-analysis found a correlational value of $r = 0.296$ for broad internalizing problems, while Stratis and Lecavalier (2015) reported $r = 0.25$ and Huang (2017) reported $r = 0.18$) with all three studies yielding a small effect size--although the current meta-analysis estimate was within rounding error of a moderate effect size value. Though the precise operationalization of "internalizing" varied across the three meta-analyses, the three resulting estimates were reasonably similar. Finally, within the teacher vs. self rater-pair, the current study yielded a correlational value of $r = 0.316$ for broad internalizing problems with Stratis and Lecavalier (2015) finding $r = 0.25$ and Huang (2017) finding $r = 0.19$. In this case, the effect size from the current meta-analysis yielded a moderate effect while the estimates from the other two meta-analyses yielded small effects. Yet, the three estimates are, again, reasonably similar to each other—especially if the range of values are viewed as sampling variation around a "true" central value. Overall, correlational agreement from across the three rater pairs in the present analysis of exclusively ASD studies yielded slightly higher, but reasonably similar estimates when compared to those derived from mixed ASD and ID (without ASD), typically developing, and diverse clinical samples. For mean difference effect size estimates, Huang's (2017) meta-analysis involving typically developing and varied clinical samples rated with the CBCL was used as the primary standard for benchmark estimates compared with those of the current meta-analysis involving exclusively ASD samples. (The meta-analysis by Stratis and Lecavlier [2015], involving both ASD and ID samples, did not include mean difference effect sizes.) Within the parent vs. self-report rater-pair, the current

study found a mean difference effect size estimate of g = 0.090 (parents reporting more internalizing problems), while Huang (2017) found the value g = - 0.21 (youth reporting more internalizing problems). These results indicate different patterns of reporting and also fall into separate effect size categories with the current study yielding a negligible effect and the Huang (2017) meta-analysis yielding a small effect. Within the parent vs. teacher rater-pair, this meta-analysis yielded a mean difference effect size of g = 0.153 and Huang (2017) yielded an estimate of g = 0.52. In both studies, parents were found to endorse more internalizing symptoms than teachers, but the current study found this effect size estimate to be below the benchmark value for a small effect, while Huang's (2017) effect size estimate fell within the medium effect range. Finally, within the teacher vs. self-report rater-pair, the current study yielded a mean difference effect size estimate of g = -0.033, while Huang (2017) found an estimate of g = -0.76. Again, both studies obtained results were in the same direction (i.e., that youth reported more internalizing symptoms than teachers), but the effect size in the current meta-analysis was clearly negligible, while the effect size in the Huang (2017) study was clearly medium and only slightly below the minimum for a large effect. Overall, the mean difference effect size estimates from the current meta-analysis involving ASD samples and Huang's (2017) meta-analysis of CBCL inter-rater findings in typically developing and various clinical samples, were very different. Huang's CBCL overall inter-rater standardized mean difference estimates were consistently larger than those obtained from the present meta-analysis involving ASD samples rated using various available rating scales.

Considerations Regarding Variation in Cross-Informant/Inter-Rater Findings

There are many possible explanations as to why the present meta-analysis found results that, in some instances, differed from expectations based on prior cross-informant ratings

research—whether inside or outside of the ASD context. First, cross-informant or inter-rater agreement is an understudied topic area both within the ASD literature and in the literature of the larger field--leading to a relatively small amount of prior literature to draw upon. Furthermore, a limited number of individual studies have achieved greater salience than others in the ASD inter-rater literature, which may have led to a biased view of what the overall research indicates--in the absence of a thorough review or meta-analysis that summarizes and synthesizes findings across the complete set of available studies. Next, there are theoretical considerations (e.g., the ABC model; De Los Reyes & Kazdin, 2005) that may readily explain variation in findings across cross-informant studies—though information on the potentially relevant variables are not typically assessed or reported in most inter-rater studies. Additionally, publication bias may have played a role in two overall inter-rater mean difference effect size estimates in the present meta-analysis (i.e., mean differences in parent vs. youth self-report ratings of anxiety and depression). Finally, a lack of sufficient statistical power for some of the analyses in the current meta-analysis could have led to Type II errors (false negatives) in some cases where statistical significance was expected.

The topic of multi-informant agreement within the population of youth with ASD is not commonly studied in the current research literature. That is, with over 5,000 articles (involving ASD) screened for this meta-analysis, only 75 included some measure of informant agreement (i.e., correlational values or/and means and standard deviations for the ratings). Because this topic is not widely researched, this state of affairs can lead to only a small number of studies representing the overall findings within a particular domain or subdomain of interest. This can make it difficult to gain a clear understanding of potential patterns in the data and to generate and test appropriate hypotheses. However, studying and summarizing the literature on this topic is

important both for planning future studies (i.e., what measures and sources/raters to use) and for understanding the body of results from already completed studies (i.e., judging how generalizable the results might be in terms of the construct, rater, or measurement method), which may potentially inform practice and/or foster theory development.

The ABC model (De Los Reyes & Kazdin, 2005) proposes an explanatory framework seeking to explain why informant discrepancies exist and incorporates contextual factors and potentially important differences among rater-pairs. Overall, this model suggests that informants have different perspectives in general and on the behavior of interest, which can lead to reasonable explanations for varying ratings of the same behavior within the same child across informants. Additionally, it indicates that youth behavior may genuinely vary across settings (e.g., home and school), further leading to a variation in report from those in different settings. It is reasonable to believe that, at times, informants may have varying perspectives on the same behavior (e.g., interpreting the same behavior to mean different things) and, at times, informants may have similar perspectives of the same behavior (e.g., a particularly salient and stable presentation of the behavior). Therefore, research studies may yield varying results on multi-informant agreement depending on the population, sample, rater types, rating situation (i.e., contextual factors as potential moderators of agreement or divergence), motivations of the raters, etc. Such conditional arrangements can be difficult to measure, account for, manipulate, etc., but reflect potentially important dimensions for future research to capture—if we want to better understand the influences upon and variation in cross-informant agreement.

Obtaining a sufficient level of statistical power, based on the number of studies included in an analysis, is needed for purposes of precision and sensitivity for detecting potential relationships or differences. For many of the moderator analyses completed in this meta-

analysis, only a small number of studies were available for inclusion (see Tables 11-14 and 23-25). Because of the small number of studies available, statistical power was likely less, sometimes considerably less, than adequate. Thus, non-significant results from these analyses should be interpreted with some caution. Ultimately, it is very reasonable to plan on conducting these moderator analyses again, in the future, when additional studies are available—and to recommend assessing the potential influence of these variables within future individual studies.

Strengths of the Present Study

This study is the first meta-analysis to investigate multi-informant agreement (parent vs. self, parent vs. teacher, and teacher vs. self) of anxiety, depression, and broad internalizing in youth with ASD using two different types of analyses (i.e., correlation and mean differences). A strength of the study was the inclusion and examination of three different rater-pairs. This allows for a broad investigation of multi-informant agreement by utilizing the most common rater-pairs that are found in research and practice. Further, this meta-analysis included specific coverage of anxiety, depression, and broad internalizing as separate constructs in order to examine similarities and differences in patterns of inter-rater agreement across these construct distinctions. Prior review studies/meta-analyses tended to treat estimates of these three constructs interchangeably as estimates of general internalizing; however, based on the variation among some results in this meta-analysis regarding these differing constructs, it seems justifiable to treat and evaluate them separately. Additionally, this meta-analysis covered both *r* and *g* effect size estimates in order to capture different dimensions of agreement (i.e., covariation around the rater means and cross-rater-paired standardized mean differences—evaluating both shared relative position in their respective distributions and group differences in absolute level of rating agreement).

Next, this meta-analysis involved a comprehensive process for screening, reviewing, and including studies in this meta-analysis. Search criteria were also broadly selected to ensure that important studies were not missed due to date of publication. This thorough method increased the likelihood that this meta-analysis included a sample of studies that is representative of the current literature.

Also, this meta-analysis, when possible, investigated potential moderator variables (e.g., youth cognitive ability, youth age, method of self-report administration, and correlation coefficient type) within each rater-pair and across each behavior construct. Specifically, this meta-analysis recognized and assessed the potential impact of self-report method/setting on effect size estimates of agreement, which is an important and unique contribution of this study. Method of self-report administration was found to significantly moderate the magnitude of the effect sizes for the correlation between parent and self-report of anxiety. This important finding has both research and clinical implications and suggests an area in need of further research exploration going forward.

Finally, the current study assessed for estimated potential bias in overall average effect size estimates, which lead to the calculation and reporting of bias-corrected effect estimates for comparison with observed estimates in the context of the overall mean effect size *g* estimates for anxiety and depression in the parent vs. self rater-pair context.

Limitations of the Present Study

There were a number of limitations to the present meta-analysis. To begin, all correlation coefficients across the available studies were treated as Pearson's *r*. This method was modeled from Stratis and Lecavlier (2015) who also treated all correlation coefficients as Pearson's *r*. This strategy which was employed to increase the number of studies available for the

correlational analyses. If the correlation coefficients were analyzed separately, it would have tripled the number of analyses required and each separate analysis would only have included a small number of studies—resulting in lower statistical power per analysis. Although correlation coefficient type was examined as a potential moderator variable and found to be non-significant and non-substantive, this pooling method is not ideal. Pearson's *r*, Spearman's *rho*, and the ICC do overlap considerably in what they account for—but they also involve potentially important differences (that may manifest in differences in their relative values depending on the conditions involved). Pearson's *r* is intended for assessing covariation around the respective means of two continuous variables (Rebekić et al., 2015), Spearman's *rho* is the equivalent of using ordinal ranks as input in the Pearson formula (Rebekić et al., 2015), and the ICC takes into account not just covariation in terms of relative distance of different means, but also in terms of absolute rating level (i.e., absolute agreement; Liu et al., 2016).

Similarly, mean comparisons between two rater types who are rating the same target, require taking into account the correlation between the rater-pairs to obtain the appropriate standard error for the effect size *g*. Not all individual studies report this correlation or even treat their mean comparisons as dependent (in fact, many meta-analyses appear to ignore this dependency issue and may treat rater means as if they are independent). As a result, when an individual mean-comparison study does not make available the correlation between rater-pairs, or other information that would allow one to derive it (e.g., reporting the standard deviation of the paired differences between raters), the correlation must be estimated. Given that the present meta-analysis involved examining inter-rater agreement from both a correlation and mean difference perspective, the present meta-analysis itself offered overall average correlation estimates for particular types of rater-pairs. Thus, when study-specific correlations were not

available to account for dependency between raters in the mean difference context, the mean correlation for that rater-pair type calculated from the present meta-analysis was used as the best estimate for purposes of estimating the standard error for that effect size *g*. Though not perfect (as an average *r* can over or under-estimate what the study-specific value would have been), this was deemed the best available option and very reasonable in terms of estimation of missing information. Because this method involves use of a very robust average *r* estimate calculated based on all available studies, any study-specific variation would likely be averaged over in the larger analysis—and some would argue that, depending on the situation, this pooled *r* estimate would likely be closer to the true population value than any single study-specific estimate.

Additionally, as mentioned before, only a small number of studies were available for some of the analyses. Specifically, there were fewer studies for some rater-pairs, the moderator analyses as a whole, and the broad internalizing construct. Overall, the greatest number of studies available for analyses were within the parent and self rater-pair and the anxiety construct. Because some analyses included as little as four studies, the average values calculated may not truly represent the topic of interest.

Another limitation of this meta-analysis was the inability to analyze all potentially relevant moderator variables. This was due to lack of information published in the studies that were included in the current meta-analysis. Ideally, more moderator variables would have been evaluated and some may have been found to account for significant true variation in effect size estimates.

Next, the range of mean FSIQ scores reported in the studies included in this meta-analysis was somewhat restricted. That is, FSIQs in the ID range (i.e., <70), were not well-represented in this meta-analysis. Researchers may shy away from utilizing self-report rating

scales with individuals functioning in the ID range because they may perceive that these youth may be unable to complete, or less able to accurately complete, a self-report rating scale due to the complex cognitive and language demand presented by self-report rating scales (Emerson, Felce, & Stancliffe, 2013). If this is true, it may explain the relative lack of available studies involving cases with IQs less than 70. With a wider range of FSIQ scores represented, it is possible that cognitive ability of youth would become a significant moderator variable. Thus, range restriction on the FSIQ variable may have attenuated the proposed relationship between cognitive ability and agreement involving self-report.

Similarly, an important consideration for the mean age variable as a potential moderator is that constructs, such as depression, may show a developmental pattern that involves different symptoms at different ages. In childhood, depression may be expressed, to some degree, through more externalizing symptoms, while as adolescence progresses, the symptoms may appear more consistent with expectations for adults (Frick et al., 1994; Ghaziuddin et al., 2002). Given that many studies involved samples that covered a broad age range (including both children and adolescents), it is possible that important developmental variation may have been averaged over in the reporting of the mean value and using this mean value to characterize the sample. In addition, the use of a single depression measure across a wide age range, covering different developmental periods, may not have included broad enough item content to capture the variation in depression symptoms across such a wide age span.

Finally, although this meta-analysis included three popular rater-pairs (i.e., parent vs. self, teacher vs. self, and parent vs. teacher), there are other rater-pairs of potential interest such as parent vs. parent, teacher vs. teacher, clinician vs. youth self-report, clinician vs. parent, or clinician vs. teacher that were not covered. Some of these rater-pairs have been examined

(Achenbach et al., 1987) in other meta-analyses that included a broader sample (i.e., not exclusively a sample of youth with ASD), indicating that it may be important to include these rater-pairs.

Implications for Future Research

As indicated in the study limitations sections, though examination of correlation coefficient type as a moderator did not reveal any significant or substantive differences, strictly speaking, there are still potentially important differences between them. From the perspective of a meta-analysis, and for more general comparative interpretation purposes across studies, having the same type of coefficient reported across studies is critical. Given that the ICC accounts for both relative covariation and more absolute (level) agreement (Liu et al., 2016), it appears to be the best suited among the three options for this purpose. However, there is also a compelling argument for inter-rater researchers to consistently report all three types of agreement coefficients. This has the advantages of allowing one to view agreement from different perspectives, provides meta-analysts with a range of effect size options from which to select, and assures comparability with prior studies that reported only one type of coefficient. Therefore, it is recommended that inter-rater researchers consistently report all three types of coefficient to assist with cross-study comparability and meta-analytic summary.

Similarly, the reporting of both correlational and mean difference findings in all inter-rater studies would be useful in capturing different aspects of inter-rater agreement (i.e., with *r* capturing covariation between raters around their respective rater means, and mean differences capturing between-group or group-level variation across rater types) in a manner that may be most directly meaningful to readers. Thus, going forward, researchers reporting inter-rater

findings are encouraged to report both correlational and mean difference results to capture both of these dimensions of agreement/disagreement.

The thoroughness of the rater-pairs, constructs, and effect size types covered revealed clear areas in need of further research to fill in gaps in terms of the number of studies available and the need for more comprehensive reporting of potential moderator variables across future individual inter-rater studies. First, the literature could clearly use more studies of parent vs. teacher and teacher vs. youth self-report ratings. More such studies would improve precision in effect size estimation and statistical power for various analyses and would allow for more expansive moderator analyses. Next, it is also important for researchers to assess for potential moderator variables when completing studies regarding multi-informant agreement of internalizing problems. As previously mentioned, very few studies included in this meta-analysis collected or reported the information needed to include the study in more than a small number of moderator analyses. In order to more thoroughly understand this research topic, analyses with strong, appropriate power are needed. It is probable that there are some factors that make agreement better among raters and knowing this information can inform the interpretation of research and practice.

The potential bias issue within the parent vs. self-report rater-pair for anxiety and depression standardized mean differences was disappointing and in need of clarification through future studies and both researcher and editorial considerations. The general pattern suggested that smaller studies were much more likely to favor a significant or even substantial parent > youth self-report rating standardized mean difference pattern, while this pattern was much less pronounced among larger studies. This leads to questions about whether editors were less likely

to accept a manuscript with non-significant mean differences between these raters or if researchers themselves were less likely to submit non-significant results for publication.

Providing youth with self-report rating scales assumes that they are able to accurately utilize a Likert scale and discriminate between categories such as “sometimes” and “often”. Making this differentiation can be challenging for adults and even more so for a child or adolescent with ASD who may have deficits in self-insight. Therefore, it could be useful to attempt to help youth understand the Likert scale rating system prior to completing a self-report assessment. In their study of factors influencing the agreement between parent and child reports of anxiety in youth with ASD, Ooi et al. (2016) taught the concept of frequency terms to the youth participants prior to them completing the rating scale. The authors did this by using an interactive computer application that visually represented the frequency of occurrence of neutral items (e.g., fruits). This study did find a significant, moderate effect for agreement ($r = 0.38$) between parents and youth. It is possible that by teaching these youth participants how to use a frequency rating scale, they were able to produce a more accurate report of their symptoms. Therefore, it may be useful for researchers to adopt a similar procedure prior to obtaining self-ratings from youth with ASD.

Finally, given the potential impact and interpretive implications of self-report ratings being impacted by completion in home vs. clinic settings and who oversees rating completion, this issue should be investigated further to more clearly understand what brings it about. For example, it is possible that parents are assisting children in completing the ratings and, thus, rendering the ratings less independent, that children are responding—when supervised by parents—in a manner more consistent with what they believe their parents want, or that children are less comfortable in the clinical setting, which impacts their ratings.

Implications for Future Practice

This meta-analysis supports that agreement among informants regarding internalizing symptoms in youth with ASD tends to be moderate at best. Therefore, it would be helpful for clinicians to attempt to gain a better understanding of the youths' symptoms through methods other than rating scales. For example, consistent with best practice regarding multiple method and multiple source assessment, diagnostic interviews and direct observation should also be utilized when possible—especially for diagnostic assessment where clinical interviews are considered essential for internalizing issues (Gray et al., 2009; Klein et al., 2005; Nardi, 2007; Silverman & Ollendick, 2005). When using these methods of assessment, clinicians should assess for daily, social, and school functioning as this information can help these professionals gain a better understanding of potential internalizing problems and the effects they may or may not have on youth functioning. Completing a diagnostic interview or a direct observation will also remove the possibility of youth informants not knowing how to rate their behaviors on a scale.

For correlation effect size estimates, findings of this meta-analysis suggest that parent vs. self agreement (medium effect) tends to be larger than teacher vs. self agreement (small effect, although non-significant) and parent vs. teacher agreement (small effect). However, even the strongest mean *r* values indicate the majority of the variance in ratings is not shared between any rater-pairs. For mean difference effect size estimates, parent vs. teacher mean differences tend to be smaller than parent vs. self or teacher vs. self mean differences. Further, in most cases, parents and teachers reported higher ratings than youth self-reporters (note: within the teacher vs. self rater-pair, youth endorsed higher ratings within the broad internalizing construct). Therefore, ratings tend to diverge across raters—which may be due to differences in perceptions

of the behavior itself, different understandings of the ordered rating categories, or real variation in behavior across context. Thus, different sources and methods are needed in a comprehensive assessment to capture different sources of variation and look for patterns of agreement and divergence.

This meta-analysis revealed two significant moderator analyses—youth age as a moderator of the correlation between parent vs. self-report of depression [with outlier removed] and method of self-report administration as a moderator of the correlation between parent vs. self-report of anxiety. First, regarding age as a moderator of the correlation between parent vs. self-report of depression, this analysis suggested that as youth age increased, the agreement between parent and self-report of depression increased. Next, the method of self-report administration (i.e., completed at home or completed in clinic) was found to moderate the correlation between parent vs. self-reported anxiety such that the correlation between parents and youth was higher (i.e., better agreement) when the assessment was completed at home. This could mean that when youth self-report ratings are completed in the home setting, they are less likely to be independent. Overall, these significant moderator variables are useful for clinicians to consider when interpreting evaluation results.

Summary and Conclusion

The present meta-analysis yielded its most significant findings within the parent vs. youth self-report rater-pair and weakest findings in relation to the teacher vs. youth self-report rater-pair. This pattern may have been related to the trend of more studies being available that examined parent vs. youth self-report ratings, somewhat fewer that examined parent vs. teacher ratings, and very few that examined teacher vs. youth self-report ratings. Within the anxiety and depression constructs for the parent vs. self rater-pair, correlation effect size estimates and mean

difference effect size estimates were all significant. These findings suggested that, although there was significant agreement within this rater-pair when parents and youth reported on anxiety and depression in youth with ASD, there were also significant observed mean differences between their ratings. The mean correlation effect size for broad internalizing within this rater-pair was also significant, while mean differences were not. Additionally, correlation effect size estimates were stronger and mean difference effect size estimates were smaller when the self-report rating scale was completed in the home setting as opposed to in a clinic. Though not clear at present, this finding could mean that there is less independence between parent and youth self-report ratings when youth ratings are completed at home with potential assistance from or influence of the parent. Further, age of youth was found to significantly moderate the relationship (correlation effect size) between parent vs. youth self-report ratings of depression—with older age leading to stronger agreement. Although the parent vs. self rater-pair involved the greatest number of significant findings, it was also the only rater-pair that showed evidence of potential publication bias. Specifically, visual inspection of the funnel plots for the standardized mean difference effect size estimates within the anxiety and depression constructs lead to utilizing the Trim and Fill method (Duval & Tweedie, 2000), which suggested that bias-adjusted effect size estimates were close to 0 (small-to-negligible and non-significant).

Differences in the number of studies available for different rater-pairs were substantive and sufficient to have influenced the pattern of results across rater-pairs. The parent vs. youth self-report rater-pair, for anxiety and depression, typically involved sufficient available studies to perform the overall mean effect size test for both *r* and *g*, conduct the bias evaluation, and to perform a significance test for each individual moderator. For internalizing, *k* was sufficient to test the significance of the overall mean effect size for *r* and arguably for *g* (which was close to

0, though based on only 4 studies), but the number of studies was insufficient to perform reasonable tests for moderators.

When considering the parent vs. teacher rater-pair, there were typically enough studies available to test the significance of the overall mean effect size estimates for *r* and *g*, and to conduct the bias evaluation. However, in most cases, the number of studies available for moderator analyses was insufficient.

In the teacher vs. youth self-report rating context, the number of studies available was generally too small for strong tests of statistical significance for overall mean effects for *r* or *g*, and for all moderator analyses—and, typically, insufficient to conduct reasonable bias evaluations. Results for the teacher vs. youth self-report overall mean effect size estimates were reported only for descriptive purposes. Despite lack of statistical significance, they do descriptively reflect the current state of the literature in terms of studies available and obtained effect size estimates. These estimates may also inform power analyses for future studies.

Moving forward, it would be helpful for those who conduct and report research concerned with cross-informant agreement to broaden the scope of their analyses and data reported when completing studies of inter-rater agreement for internalizing problems in youth with ASD. Specifically, researchers should (a) report multiple correlation coefficient types, (b) include both inter-rater correlation and mean difference effect sizes, and (c) collect and report detailed information that could be used in moderator analyses (i.e., age of youth, cognitive ability of youth, method of self-report administration, language ability of youth, parent depression, etc.). These results have important implications for both researchers and clinicians who work with youth with ASD and comorbid internalizing conditions.

The imperfect magnitude of inter-rater correlations, variability in mean inter-rater correlations across types of rater-pairs, and variability across studies in cross-rater mean differences all suggest that reliance on a single rater type to measure an internalizing outcome among youth with ASD could yield results that do not generalize well across those for other rater types. Thus, in the absence of more objective indicators of internalizing issues in youth with ASD, researchers are urged to regularly practice multi-operationalization of internalizing constructs in their studies and clinicians should emphasize best-practice use of multiple methods and multiple sources in comprehensively assessing internalizing symptoms in ASD—using strategies that complement each other and address each other's weaknesses. Finally, the more negative outcomes associated with internalizing issues in ASD highlight the importance of identification and treatment. Screening is important; however, reliance on a single rater for screening purposes likely invites the greater possibility of false negative detection errors (i.e., failure to identify internalizing problems that are actually present)—a much more problematic error than a false positive screening result.

Table 1.

Prior Meta-Analyses

Author(s)	**Population**	**Construct**	**Assessment**	**Informant(s)**	**Results**
Achenbach et al., 1987 (*K* = 119)	Diverse sample of youth	Behavior and Emotional Problems	Rating scales	Youth-parent (YP) Youth-teacher (YT) Parent-teacher (PT)	YP total behavior: *r* = .25 YT total behavior: *r* = .20 PT total behavior: *r* = .27
Achenbach et al., 2005 (*K* = 108)	Adults	Substance Use, Internalizing Problems, and Externalizing Problems	Clinical interviews and questionnaires	Self-informant	Internalizing problems: *r* = .43 Externalizing problems: *r* = .44
Stratis & Lecavalier, 2015 (*K* = 49)	Youth with ASD or ID	Externalizing Problems, Internalizing Problems, and Social Skills	Behavioral rating scales	Youth-parent (YP) Youth-teacher (YT) Parent-teacher (PT)	<u>Internalizing Problems</u> YP: *r* = .42 YT: *r* = .25 PT: *r* = .25 <u>Externalizing Problems</u> YP: *r* = .44 YT: *r* = .34 PT: *r* = .38
Huang, 2017 (*K* = 169)	Non-ASD youth	Behavior and Emotional Problems	Child Behavior Checklist (CBCL)	Youth-parent (YP) Youth-teacher (YT) Parent-teacher (PT)	<u>Mean-level Agreement</u> YP internalizing: *g* = -.21 YT internalizing: *g* = -.76 PT internalizing: *g* = .52 <u>Rank-order Agreement</u> YP internalizing: *r* = .33 YT internalizing: *r* = .19 PT internalizing: *r* = .18

Table 1 (cont'd)

Author(s)	Population	Construct	Assessment	Informant(s)	Results
Van Steensel & Heeman, 2017 (*K* = 83)	Youth with ASD, youth without ASD, and clinically referred youth	Anxiety	Anxiety inventories	Youth-parent (YP)	ASD vs. TD Fixed model: *d* = .78 Random model: *d* = .97 ASD vs. clinically referred Fixed model: *d* = .23 Random model: *d* = .12
Hudson, Hall, & Harkness, 2018 (*K* = 66)	Children, adolescents, and adults with ASD	Depression	Clinical interviews and questionnaires	Parent and individual self-report	Children (18 and under) Current prevalence: 10.6% Lifetime prevalence: 7.7% Adults (18 and over) Current prevalence: 19.4% Lifetime prevalence: 40.2%
Hollocks et al., 2018 (*K* = 35)	Adults with ASD	Anxiety and Depression	Clinical interviews and questionnaires	Adult self-report	Current Prevalence Any anxiety disorder: 27% Depression: 23% Lifetime Prevalence Any anxiety disorder: 42% Depression: 37%

Table 2.

Study-by-Study Correlation Between Parent-rated and Self-rated Anxiety

Study	Measure(s)	*N* pair	Correlation Coefficient	Correlation Value	Z-Value	*p*-Value	95% Confidence Interval
Bermudez et al., 2015	SCARED	38	Pearson's *r*	0.319	1.955	0.051	-0.001, 0.580
Bitsika et al., 2019	CASI-4 (GAD)	150	Pearson's *r*	0.570	7.851	< 0.001	0.451, 0.669
Blakeley-Smith et al., 2012	SCARED	63	ICC	0.520	4.464	< 0.001	0.312, 0.680
Chow, 2008	BASC-2 PRS and SRP	32	Pearson's *r*	-0.022	-0.118	0.906	-0.368, 0.329
Farrugia and Hudson, 2006	SCAS	29	Pearson's *r*	0.697	4.393	< 0.001	0.444, 0.847
Freeman, 2009	Revised Children's Manifest Anxiety Scale (RCMAS) vs. BASC-2	61	Pearson's *r*	0.550	4.709	< 0.001	0.346, 0.704
Hallett et al., 2013	RCADS	79	Pearson's *r*	0.490	4.673	< 0.001	0.302, 0.642
Hurtig et al., 2009	ASEBA CBCL and YSR	45	Spearman's *rho*	0.290	1.935	0.053	-0.004, 0.538
Jepsen et al., 2012	BASC-2 PRS and BASC-2 SRP	40	Pearson's *r*	0.134	0.820	0.412	-0.185, 0.428
Kaat, 2014	ASEBA CBCL and YSR	44	ICC	0.529	3.770	< 0.001	0.275, 0.714
Lohr et al., 2017	RCADS	41	Pearson's *r*	0.230	1.444	0.149	-0.084, 0.502
Lopata et al., 2010	SCARED	73	Pearson's *r*	0.430	3.848	< 0.001	0.222, 0.601

Table 2 (cont'd)

Magiati, 2014	SCAS	38	ICC	0.690	5.017	< 0.001	0.475, 0.827
Mertens et al., 2017	SCARED	22	Spearman's *rho*	0.270	1.207	0.228	-0.171, 0.621
Ooi et al., 2016	SCAS	70	ICC	0.380	3.275	0.001	0.159, 0.565
Ozsivadjian et al., 2014	SCAS	30	ICC	0.590	3.521	< 0.001	0.292, 0.784
Pisula et al., 2017	ASEBA CBCL and YSR	35	Pearson's *r*	0.420	2.533	0.011	0.101, 0.661
Rosen and Lerner, 2018	MASC	51	Pearson's *r*	0.050	0.347	0.729	-0.229, 0.321
Rosenberg, 2016	BASC-2 PRS and SRP	20	Pearson's *r*	0.590	2.794	0.005	0.200, 0.819
Rump, 2012	SCARED	19	Pearson's *r*	0.190	0.769	0.442	-0.289, 0.593
Schiltz et al., 2018	ASEBA CBCL and YSR	53	Pearson's *r*	0.370	2.747	0.006	0.111, 0.582
Schwartz, 2010	BASC-2 PRS and SRP	30	Pearson's *r*	0.210	1.108	0.268	-0.163, 0.530
Sterling et al., 2015	RCADS and ASEBA CBCL	67	Pearson's *r*	0.260	2.129	0.033	0.021, 0.471
Taylor et al., 2018	BASC-2 PRS and SRP	44	Pearson's *r*	0.198	1.285	0.199	-0.105, 0.467
Whitehead, 2005	BASC PRS and SRP	20	Pearson's *r*	0.480	2.156	0.031	0.048, 0.761
Overall				**0.399**	**9.294**	**< 0.001**	**0.321, 0.471**

Note. All *p* values reflect two-tailed probabilities.

Table 3.

Study-by-Study Correlation Between Parent-rated and Self-rated Depression

Study	**Measure(s)**	***N* pair**	**Correlation Coefficient**	**Correlation Value**	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Bohnert et al., 2016	ASEBA CBCL and YSR	127	Pearson's *r*	0.610	7.894	< 0.001	0.488, 0.709
Chow, 2008	BASC-2 PRS and CDI	32	Pearson's *r*	0.292	1.620	0.105	-0.063, 0.581
Freeman, 2009	BASC-2 PRS and CDI	61	Pearson's *r*	0.480	3.983	< 0.001	0.260, 0.653
Hurtig et al., 2009	ASEBA CBCL and YSR	45	Spearman's *rho*	0.290	1.935	0.053	-0.004, 0.538
Hammond and Hammond, 2014	Adolescent Symptom Inventory (ASI-4) and Youth Inventory (YI-4)	12	Pearson's *r*	0.610	2.127	0.033	0.056, 0.877
Jepsen et al., 2012	ASEBA CBCL and YSR	44	ICC	0.524	3.726	< 0.001	0.269, 0.710
Kaat, 2014	RCADS	41	Pearson's *r*	0.360	2.323	0.020	0.059, 0.601
Lee, 2009	BASC-2 PRS and SRP	30	Pearson's *r*	0.280	1.495	0.135	-0.089, 0.582
Lopata et al., 2010	BASC-2 PRS and SRP	40	Pearson's *r*	0.315	1.984	0.047	0.004, 0.571
Ozsivadjian et al., 2014	CDI	30	ICC	0.620	3.767	< 0.001	0.334, 0.801
Pisula et al., 2017	ASEBA CBCL and YSR	35	Pearson's *r*	0.420	2.533	0.011	0.101, 0.661
Rosen and Lerner, 2018	BASC-2	51	Pearson's *r*	0.140	0.976	0.329	-0.141, 0.400

Table 3 (cont'd)

Rosenberg, 2016	BASC-2 PRS and CDI	20	Pearson's *r*	0.450	1.998	0.046	0.009, 0.744
Rump, 2012	CDI	19	Pearson's *r*	-0.160	-0.646	0.519	-0.573, 0.317
Taylor et al., 2018	BASC-2 PRS and SRP	44	Pearson's *r*	0.368	2.472	0.013	0.080, 0.599
Vickerstaff et al., 2007	BASC PRS and SRP	22	Pearson's *r*	0.670	3.534	< 0.001	0.346, 0.851
Overall				**0.412**	**7.486**	**< 0.001**	**0.313, 0.503**

Note. All *p* values reflect two-tailed probabilities.

Table 4.

Study-by-Study Correlation Between Parent-rated and Self-rated Broad Internalizing

Study	**Measure(s)**	***N* pair**	**Correlation Coefficient**	**Correlation Value**	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Bitsika et al., 2019	CASI-4	150	Pearson's *r*	0.606	8.518	< 0.001	0.242, 0.587
Hurtig et al., 2009	ASEBA CBCL and YSR	45	Spearman's *rho*	0.270	1.794	0.073	-0.026, 0.522
Jepsen et al., 2012	ASEBA CBCL and YSR	44	ICC	0.564	4.090	< 0.001	0.321, 0.737
Kaat and Lecavalier, 2015	RCADS	46	ICC	0.250	1.675	0.094	-0.043, 0.504
Pisula et al., 2017	ASEBA CBCL and YSR	35	Pearson's *r*	0.310	1.813	0.070	-0.026, 0.583
Overall				**0.430**	**4.227**	**< 0.001**	**0.242, 0.587**

Note. All *p* values reflect two-tailed probabilities.

Table 5.

Study-by-Study Correlation Between Parent-rated and Teacher-rated Anxiety

Study	**Measure(s)**	***N* pair**	**Correlation Coefficient**	**Correlation Value**	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Adams et al., 2018	The Anxiety Scale for Children with Autism Spectrum Disorder	92	Spearman's *rho*	0.390	3.885	< 0.001	0.201, 0.551
Hurtig et al., 2009	ASEBA CBCL and TRF	22	Spearman's *rho*	0.190	0.838	0.402	-0.252, 0.566
Jepsen et al., 2012	ASEBA CBCL and TRF	36	ICC	0.286	1.578	0.115	-0.066, 0.548
Kanne et al., 2009	ASEBA CBCL and TRF	177	Pearson's *r*	0.140	1.859	0.063	-0.008, 0.282
Lane et al., 2013	ASEBA CBCL and TRF	39	Pearson's *r*	0.260	1.597	0.110	-0.060, 0.532
McDonald et al., 2016	BASC-2 PRS and TRS	118	Pearson's *r*	0.340	3.797	< 0.001	0.170, 0.491
Ung et al., 2017	ASEBA CBCL and TRF	32	ICC	0.410	2.346	0.019	0.072, 0.664
Overall				**0.273**	**5.908**	**< 0.001**	**0.185, 0.356**

Note. All *p* values reflect two-tailed probabilities.

Table 6.

Study-by-Study Correlation Between Parent-rated and Teacher-rated Depression

Study	**Measure(s)**	***N* pair**	**Correlation Coefficient**	**Correlation Value**	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Hurtig et al., 2009	ASEBA CBCL and TRF	22	Spearman's *rho*	0.310	1.397	0.162	-0.128, 0.647
Jepsen et al., 2012	ASEBA CBCL and TRF	36	ICC	0.351	2.106	0.035	0.025, 0.609
Kanne et al., 2009	ASEBA CBCL and TRF	177	Pearson's *r*	0.080	1.058	0.290	-0.068, 0.225
Lane et al., 2013	ASEBA CBCL and TRF	39	Pearson's *r*	0.450	2.908	0.004	0.157, 0.670
McDonald et al., 2016	BASC-2 PRS and TRS	118	Pearson's *r*	0.300	3.319	0.001	0.126, 0.456
Ung et al., 2017	ASEBA CBCL and TRF	32	ICC	0.330	1.846	0.065	-0.021, 0.609
Vickerstaff et al., 2007	BASC PRS and TRF	22	Pearson's *r*	0.220	0.975	0.330	-0.222, 0.587
Overall				**0.256**	**4.222**	**< 0.001**	**0.140, 0.366**

Note. All *p* values reflect two-tailed probabilities.

Table 7.

Study-by-Study Correlation Between Parent-rated and Teacher-rated Broad Internalizing

Study	**Measure(s)**	***N* pair**	**Correlation Coefficient**	**Correlation Value**	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Connolly, 2012	ASEBA CBCL and TRF	71	Pearson's *r*	0.235	1.975	0.048	0.002, 0.444
Dauterman, 2017	BASC-2 PRS and TRS	70	Pearson's *r*	0.600	5.674	< 0.001	0.425, 0.732
Hurtig et al., 2009	ASEBA CBCL and TRF	22	Spearman's *rho*	0.050	0.218	0.827	-0.380, 0.462
Jepsen et al., 2012	ASEBA CBCL and TRF	36	ICC	0.212	1.237	0.216	-0.125, 0.505
Lane et al., 2013	ASEBA CBCL and TRF	39	Pearson's *r*	0.300	1.857	0.063	-0.017, 0.562
McDonald et al., 2016	BASC-2 PRS and TRS	118	Pearson's *r*	0.280	3.085	0.002	0.105, 0.439
Peterson, 2017	BASC-2 PRS and TRS	26	Pearson's *r*	0.580	3.177	0.001	0.248, 0.790
Rodriguez, 2017	ASEBA CBCL and TRF	166	Pearson's *r*	0.120	1.539	0.124	-0.033, 0.267
Ung et al., 2017	ASEBA CBCL and TRF	32	ICC	0.180	0.980	0.327	-0.180, 0.497
Overall				**0.296**	**4.131**	**< 0.001**	**0.159, 0.422**

Note. All *p* values reflect two-tailed probabilities.

Table 8.

Study-by-Study Correlation Between Teacher-rated and Self-rated Anxiety

Study	**Measure(s)**	***N* pair**	**Correlation Coefficient**	**Correlation Value**	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Hurtig et al., 2009	ASEBA CBCL and YSR	23	Spearman's *rho*	0.340	1.584	0.113	-0.084, 0.660
Jepsen et al., 2012	ASEBA CBCL and YSR	36	ICC	0.158	0.915	0.360	-0.180, 0.464
Overall				**0.229**	**1.695**	**0.090**	**-0.036, 0.464**

Note. All *p* values reflect two-tailed probabilities.

Table 9.

Study-by-Study Correlation Between Teacher-rated and Self-rated Depression

Study	**Measure(s)**	***N* pair**	**Correlation Coefficient**	**Correlation Value**	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Hurtig et al., 2009	ASEBA CBCL and YSR	23	Spearman's *rho*	0.660	3.546	< 0.001	0.340, 0.843
Jepsen et al., 2012	ASEBA CBCL and YSR	36	ICC	0.091	0.524	0.600	-0.245, 0.407
Vickerstaff et al., 2007	BASC TRF and SRP	22	Pearson's *r*	0.220	0.975	0.330	-0.222, 0.587
Overall				**0.342**	**1.660**	**0.097**	**-0.064, 0.651**

Note. All *p* values reflect two-tailed probabilities.

Table 10.

Study-by-Study Correlation Between Teacher-rated and Self-rated Broad Internalizing

Study	**Measure(s)**	***N* pair**	**Correlation Coefficient**	**Correlation Value**	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Hurtig et al., 2009	ASEBA CBCL and YSR	23	Spearman's *rho*	0.560	2.830	0.005	0.192, 0.790
Jepsen et al., 2012	ASEBA CBCL and YSR	36	ICC	0.056	0.322	0.747	-0.278, 0.378
Overall				**0.316**	**1.137**	**0.255**	**-0.233, 0.712**

Note. All *p* values reflect two-tailed probabilities.

Table 11.

Information for Studies Included in Analyses for Youth Cognitive Ability as a Continuous Moderator for Mean Correlational Values

Study	Rater-pair	Construct	*M* FSIQ Score	Correlation Value	*N* pair
Bitsika et al., 2019	Parent vs. Self	Anxiety	94.98	0.570	150
Chow, 2008	Parent vs. Self	Anxiety	107.66	-0.022	32
Hallett et al., 2013	Parent vs. Self	Anxiety	88.07	0.490	79
Jepsen et al., 2012	Parent vs. Self	Anxiety	91.13	0.134	40
Kaat, 2014	Parent vs. Self	Anxiety	90.70	0.529	44
Lopata et al., 2010	Parent vs. Self	Anxiety	110.14	0.430	73
Ozsivadjian et al., 2014	Parent vs. Self	Anxiety	94.90	0.590	30
Pisula et al., 2017	Parent vs. Self	Anxiety	103.94	0.420	35
Rosen and Lerner, 2018	Parent vs. Self	Anxiety	102.82	0.050	51
Rump, 2012	Parent vs. Self	Anxiety	106.00	0.190	19
Schiltz et al., 2018	Parent vs. Self	Anxiety	104.96	0.370	53
Sterling et al., 2015	Parent vs. Self	Anxiety	87.78	0.260	67
Taylor et al., 2018	Parent vs. Self	Anxiety	101.66	0.198	44
Bohnert et al., 2016	Parent vs. Self	Depression	104.76	0.610	127
Chow, 2008	Parent vs. Self	Depression	107.66	0.292	32
Jepsen et al., 2012	Parent vs. Self	Depression	91.13	0.524	44
Kaat, 2014	Parent vs. Self	Depression	90.70	0.360	41
Lee et al., 2009	Parent vs. Self	Depression	104.00	0.280	30
Lopata et al., 2010	Parent vs. Self	Depression	110.14	0.315	40
Ozsivadjian et al., 2014	Parent vs. Self	Depression	94.90	0.620	30

Table 11 (cont'd)

Pisula et al., 2017	Parent vs. Self	Depression	103.94	0.420	35
Rosen and Lerner, 2018	Parent vs. Self	Depression	102.82	0.140	51
Rump, 2012	Parent vs. Self	Depression	106.00	-0.160	19
Taylor et al., 2018	Parent vs. Self	Depression	101.66	0.368	44
Vickerstaff et al., 2007	Parent vs. Self	Depression	105.41	0.670	22
Bitsika et al., 2019	Parent vs. Self	Internalizing	94.90	0.606	150
Jepsen et al., 2012	Parent vs. Self	Internalizing	91.13	0.564	44
Kaat and Lecavalier, 2015	Parent vs. Self	Internalizing	90.70	0.250	46
Pisula et al., 2017	Parent vs. Self	Internalizing	103.94	0.310	35
Jepsen et al., 2012	Parent vs. Teacher	Anxiety	91.13	0.286	36
McDonald et al., 2016	Parent vs. Teacher	Anxiety	103.12	0.340	118
Ung et al., 2017	Parent vs. Teacher	Anxiety	81.30	0.410	32
Jepsen et al., 2012	Parent vs. Teacher	Depression	91.13	0.351	36
McDonald et al., 2016	Parent vs. Teacher	Depression	103.12	0.300	118
Ung et al., 2017	Parent vs. Teacher	Depression	81.30	0.330	32
Vickerstaff et al., 2007	Parent vs. Teacher	Depression	105.41	0.220	22
Jepsen et al., 2012	Parent vs. Teacher	Internalizing	91.13	0.212	36
McDonald et al., 2016	Parent vs. Teacher	Internalizing	103.12	0.280	118
Rodriguez, 2017	Parent vs. Teacher	Internalizing	93.86	0.120	166
Ung et al., 2017	Parent vs. Teacher	Internalizing	81.30	0.180	32

Table 11 (cont'd)

Jepsen et al., 2012	Teacher vs. Self	Anxiety	91.13	0.158	36
Jepsen et al., 2012	Teacher vs. Self	Depression	91.13	0.091	36
Vickerstaff et al., 2007	Teacher vs. Self	Depression	105.41	0.220	22
Jepsen et al., 2012	Teacher vs. Self	Internalizing	91.13	0.056	36

Table 12.

Information for Studies Included in Analyses for Youth Mean Age in Years as a Continuous Moderator for Mean Correlational Values

Study	**Rater-pair**	**Construct**	***M* Age (years)**	**Correlation Value**	***N* pair**
Bermudez et al., 2015	Parent vs. Self	Anxiety	12.15	0.319	38
Bitsika et al., 2019	Parent vs. Self	Anxiety	11.20	0.570	150
Blakeley-Smith et al., 2012	Parent vs. Self	Anxiety	10.10	0.520	63
Chow, 2008	Parent vs. Self	Anxiety	10.28	-0.022	32
Farrugia and Hudson, 2006	Parent vs. Self	Anxiety	13.80	0.697	29
Freeman, 2009	Parent vs. Self	Anxiety	11.96	0.550	61
Hallett et al., 2013	Parent vs. Self	Anxiety	13.50	0.490	79
Hurtig et al., 2009	Parent vs. Self	Anxiety	13.00	0.290	45
Jepsen et al., 2012	Parent vs. Self	Anxiety	15.06	0.134	40
Kaat, 2014	Parent vs. Self	Anxiety	12.40	0.529	44
Lohr et al., 2017	Parent vs. Self	Anxiety	12.90	0.230	41
Lopata et al., 2010	Parent vs. Self	Anxiety	9.75	0.430	73
Magiati et al., 2014	Parent vs. Self	Anxiety	12.10	0.690	38
Mertens et al., 2017	Parent vs. Self	Anxiety	13.80	0.270	22
Ooi et al., 2016	Parent vs. Self	Anxiety	11.21	0.380	70
Ozsivadjian et al., 2014	Parent vs. Self	Anxiety	13.00	0.590	30
Pisula et al., 2017	Parent vs. Self	Anxiety	13.54	0.420	35
Rosen and Lerner, 2018	Parent vs. Self	Anxiety	12.15	0.050	51
Rump, 2012	Parent vs. Self	Anxiety	14.68	0.190	19

Table 12 (cont'd)

Schiltz et al., 2018	Parent vs. Self	Anxiety	13.45	0.370	53
Schwartz, 2010	Parent vs. Self	Anxiety	13.70	0.210	30
Sterling et al., 2015	Parent vs. Self	Anxiety	12.25	0.260	67
Taylor et al., 2018	Parent vs. Self	Anxiety	10.33	0.198	44
Whitehead, 2005	Parent vs. Self	Anxiety	14.85	0.480	20
Bohnert et al., 2016	Parent vs. Self	Depression	14	0.610	127
Chow, 2008	Parent vs. Self	Depression	10.3	0.292	32
Freeman, 2009	Parent vs. Self	Depression	13	0.480	61
Hammond and Hoffman, 2014	Parent vs. Self	Depression	14	0.610	12
Hurtig et al., 2009	Parent vs. Self	Depression	13.00	0.524	45
Jepsen et al., 2012	Parent vs. Self	Depression	15.06	0.360	44
Kaat, 2014	Parent vs. Self	Depression	12.40	0.280	41
Lee et al., 2009	Parent vs. Self	Depression	9.3	0.315	30
Lopata et al., 2010	Parent vs. Self	Depression	9.8	0.620	40
Ozsivadjian et al., 2014	Parent vs. Self	Depression	13.00	0.420	30
Pisula et al., 2017	Parent vs. Self	Depression	13.54	0.140	35
Rosen and Lerner, 2018	Parent vs. Self	Depression	12.15	0.610	51
Rump, 2012	Parent vs. Self	Depression	14.68	-0.160	19
Taylor et al., 2018	Parent vs. Self	Depression	10.33	0.368	44
Vickerstaff et al., 2007	Parent vs. Self	Depression	11.86	0.670	22

Table 12 (cont'd)

Bitsika et al., 2019	Parent vs. Self	Internalizing	11.20	0.606	150
Hurtig et al., 2009	Parent vs. Self	Internalizing	13.00	0.270	45
Jepsen et al., 2012	Parent vs. Self	Internalizing	15.06	0.564	44
Kaat and Lecavalier, 2015	Parent vs. Self	Internalizing	12.40	0.250	46
Pisula et al., 2017	Parent vs. Self	Internalizing	13.54	0.310	35
Hurtig et al., 2009	Parent vs. Teacher	Anxiety	13.00	0.190	22
Jepsen et al., 2012	Parent vs. Teacher	Anxiety	15.06	0.286	36
Kanne et al., 2009	Parent vs. Teacher	Anxiety	7.30	0.140	177
Lane et al., 2013	Parent vs. Teacher	Anxiety	4.20	0.260	39
McDonald et al., 2016	Parent vs. Teacher	Anxiety	8.74	0.340	118
Ung et al., 2017	Parent vs. Teacher	Anxiety	7.47	0.410	32
Hurtig et al., 2009	Parent vs. Teacher	Depression	13.00	0.310	22
Jepsen et al., 2012	Parent vs. Teacher	Depression	15.06	0.351	36
Kanne et al., 2009	Parent vs. Teacher	Depression	7.30	0.080	177
Lane et al., 2013	Parent vs. Teacher	Depression	4.20	0.450	39
McDonald et al., 2016	Parent vs. Teacher	Depression	8.74	0.300	118
Ung et al., 2017	Parent vs. Teacher	Depression	7.47	0.330	32
Vickerstaff et al., 2007	Parent vs. Teacher	Depression	11.86	0.220	22
Connolly, 2012	Parent vs. Teacher	Internalizing	9.30	0.235	71
Dauterman, 2017	Parent vs. Teacher	Internalizing	4.80	0.600	70

Table 12 (cont'd)

Hurtig et al., 2009	Parent vs. Teacher	Internalizing	13.00	0.050	22
Jepsen et al., 2012	Parent vs. Teacher	Internalizing	15.06	0.212	36
Lane et al., 2013	Parent vs. Teacher	Internalizing	4.20	0.300	39
McDonald et al., 2016	Parent vs. Teacher	Internalizing	8.74	0.280	118
Peterson, 2017	Parent vs. Teacher	Internalizing	4.89	0.580	26
Rodriguez, 2017	Parent vs. Teacher	Internalizing	5.10	0.120	166
Ung et al., 2017	Parent vs. Teacher	Internalizing	7.47	0.180	32
Hurtig et al., 2009	Teacher vs. Self	Anxiety	13.00	0.340	232
Jepsen et al., 2012	Teacher vs. Self	Anxiety	15.06	0.158	36
Hurtig et al., 2009	Teacher vs. Self	Depression	13.00	0.660	23
Jepsen et al., 2012	Teacher vs. Self	Depression	15.06	0.091	36
Vickerstaff et al., 2007	Teacher vs. Self	Depression	11.86	0.220	22
Hurtig et al., 2009	Teacher vs. Self	Internalizing	13.00	0.560	23
Jepsen et al., 2012	Teacher vs. Self	Internalizing	15.06	0.056	36

Table 13.

Information Regarding Studies Included in Analyses for Method of Self-Report Administration as a Moderator for Mean Correlational Values

Study	Rater-pair	Construct	Method of Self-report Administration	Correlation Value	*N* pair
Bitsika et al., 2019	Parent vs. Self	Anxiety	Assessment completed at home	0.570	150
Freeman, 2009	Parent vs. Self	Anxiety	Assessment completed at home	0.550	61
Hallett et al., 2013	Parent vs. Self	Anxiety	Assessment completed at home	0.490	79
Jepsen et al., 2012	Parent vs. Self	Anxiety	Assessment completed at home	0.134	40
Magiati et al., 2014	Parent vs. Self	Anxiety	Assessment completed at home	0.690	38
Mertens et al., 2017	Parent vs. Self	Anxiety	Assessment completed in clinic	0.270	22
Ooi et al., 2016	Parent vs. Self	Anxiety	Assessment completed in clinic	0.380	70
Pisula et al., 2017	Parent vs. Self	Anxiety	Assessment completed in clinic	0.420	35
Schiltz et al., 2018	Parent vs. Self	Anxiety	Assessment completed in clinic	0.370	53
Schwartz, 2010	Parent vs. Self	Anxiety	Assessment completed in clinic	0.210	30
Blakeley-Smith et al., 2012	Parent vs. Self	Anxiety	Assessment read to child in clinic	0.520	63
Chow, 2008	Parent vs. Self	Anxiety	Assessment read to child in clinic	-0.022	32
Lopata et al., 2010	Parent vs. Self	Anxiety	Assessment read to child in clinic	0.430	73
Freeman, 2009	Parent vs. Self	Depression	Assessment completed at home	0.480	61
Jepsen et al., 2012	Parent vs. Self	Depression	Assessment completed at home	0.524	44
Hammond and Hoffman, 2014	Parent vs. Self	Depression	Assessment completed in clinic	0.610	12

Table 13 (cont'd)

Pisula et al., 2017	Parent vs. Self	Depression	Assessment completed in clinic	0.420	35
Chow, 2008	Parent vs. Self	Depression	Assessment read to child in clinic	0.292	32
Lee et al., 2009	Parent vs. Self	Depression	Assessment read to child in clinic	0.280	30
Lopata et al., 2010	Parent vs. Self	Depression	Assessment read to child in clinic	0.315	40
Bitsika et al., 2019	Parent vs Self	Internalizing	Assessment completed at home	0.606	150
Jepsen et al., 2012	Parent vs Self	Internalizing	Assessment completed at home	0.564	44
Pisula et al., 2017	Parent vs Self	Internalizing	Assessment completed in clinic	0.310	35
Kaat and Lecavalier, 2015	Parent vs Self	Internalizing	Assessment read to child in clinic	0.250	46

Table 14.

Study-by-Study Hedges g Values Between Parent-rated and Self-rated Anxiety

Study	*N* pair	Hedges *g*	Z-Value	*p*-Value	95% Confidence Interval
Barnhill et al., 2000	20	0.990	3.399	0.001	0.419, 1.561
Bellini, 2004	41	0.689	3.672	< 0.001	0.321, 1.056
Bitsika and Sharpley, 2015	139	0.279	2.963	0.003	0.095, 0.464
Bitsika et al., 2019	150	-0.125	-1.651	0.099	-0.273, 0.023
Bitsika et al., 2014	32	0.188	0.984	0.325	-0.186, 0.562
Blakeley-Smith et al., 2012	63	0.382	3.023	0.003	0.134, 0.630
Boulter et al., 2014	170	0.071	0.841	0.400	-0.094, 0.235
Carruthers et al., 2018	38	0.137	0.784	0.433	-0.206, 0.480
Chalfant et al., 2007	28	0.217	1.065	0.287	-0.182, 0.617
Chiu et al., 2016	28	0.355	1.708	0.088	-0.052, 0.763
Chow, 2008	32	0.265	1.055	0.292	-0.227, 0.757
Clarke et al., 2017	14	0.064	0.231	0.817	-0.477, 0.605
Conaughton et al., 2017	21	-0.667	-2.603	0.009	-1.170, -0.165
Drmic et al., 2017	35	-0.279	-1.511	0.131	-0.642, 0.083
Elzinga, 2015	26	1.149	5.042	< 0.001	0.702, 1.596
Foley Nicpon et al., 2010	25	0.460	2.053	0.040	0.021, 0.899
Freeman, 2009	61	0.314	2.553	0.011	0.073, 0.555
Hallett et al., 2013	79	-0.050	-0.445	0.657	-0.271, 0.171
Hammond and Hoffman, 2014	10	2.585	3.648	< 0.001	1.196, 3.974
Hollocks et al., 2013	38	0.126	0.721	0.471	-0.217, 0.469
Jepsen et al., 2012	44	0.233	1.599	0.110	-0.053, 0.519

Table 14 (cont'd)

Joyce et al., 2017	13	0.001	0.005	0.996	-0.557, 0.559
Kaat, 2014	43	0.196	1.051	0.293	-0.170, 0.561
Keith et al., 2018	26	-0.366	-1.695	0.090	-0.789, 0.057
Lopata et al., 2010	40	0.579	2.619	0.009	0.146, 1.012
Luxford et al., 2017	18	0.410	1.589	0.112	-0.096, 0.916
Magiati et al., 2014	38	-0.426	-3.253	0.001	-0.682, -0.169
Mertens et al., 2017	22	0.083	0.332	0.740	-0.405, 0.570
Neil et al., 2019	19	-0.086	-0.357	0.721	-0.559, 0.387
Ooi et al., 2008	6	-0.993	-2.023	0.042	-1.955, -0.031
Ooi et al., 2016	70	-0.608	-4.234	< 0.001	-0.889, -0.326
Reaven et al., 2009	10	0.472	1.399	0.162	-0.189, 1.133
Rodgers et al., 2016	157	-0.078	-0.895	0.371	-0.249, 0.093
Rosenberg, 2016	20	0.304	1.526	0.127	-0.086, 0.694
Rosen and Lerner, 2018	51	0.580	2.815	0.005	0.176, 0.983
Rump, 2012	19	-0.572	-1.883	0.060	-1.167, 0.023
Sharpley et al., 2015	16	0.090	0.345	0.730	-0.421, 0.601
Sterling et al., 2015	19	-0.190	-0.781	0.435	-0.667, 0.287
Stern et al., 2014	119	0.392	3.778	< 0.001	0.189, 0.595
Storch, 2015	16	1.809	4.143	< 0.001	0.953, 2.665
Taylor et al., 2018	44	0.665	3.197	0.001	0.257, 1.072
Van Schalkwyk et al., 2018	35	-0.091	-0.499	0.618	-0.447, 0.265
Whitehead, 2005	20	1.537	4.648	< 0.001	2.185, 4.648
Wijnhoven et al., 2018	168	0.002	0.028	0.978	-0.163, 0.167
Wood et al., 2015	14	2.220	4.140	< 0.001	1.169, 3.271

Table 14 (cont'd)

Wood et al., 2009	14	0.810	2.509	0.012	0.177, 1.443
Overall		**0.220**	**3.661**	**< 0.001**	**0.102, 0.337**

Note. All *p* values reflect two-tailed probabilities.
Note. Positive Hedges *g* values indicate parent ratings were higher than self-ratings; negative Hedges values indicate self-ratings were higher than parent ratings.

Table 15.

Study-by-Study Hedges g Values Between Parent-rated and Self-rated Depression

Study	***N* pair**	**Hedges *g***	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Barnhill et al., 2000	20	1.636	4.392	< 0.001	0.906, 2.366
Bitsika et al., 2015	150	-0.189	-2.077	0.038	-0.367, -0.011
Chow, 2008	32	0.904	3.683	< 0.001	0.423, 0.423
Foley Nicpon et al., 2010	25	0.973	3.695	< 0.001	0.457, 1.489
Freeman, 2009	61	0.963	6.148	< 0.001	0.656, 1.270
Hammond and Hoffman, 2014	12	1.738	4.423	< 0.001	0.968, 2.509
Hollocks et al., 2013	38	0.983	4.553	< 0.001	0.560, 1.406
Jepsen et al., 2012	44	0.009	0.061	0.951	-0.274, 0.292
Kaat, 2014	41	0.257	1.457	0.145	-0.089, 0.602
Lee, 2009	30	0.863	3.428	0.001	0.370, 1.357
Lopata et al., 2010	40	1.152	4.883	< 0.001	0.690, 1.614
Richdale and Baglin, 2015	17	1.001	3.139	0.002	0.376, 1.626
Rosenberg, 2016	20	0.593	2.415	0.016	0.112, 1.075
Rosen and Lerner, 2018	51	1.053	4.643	< 0.001	0.608, 1.497
Rump, 2012	19	0.773	2.007	0.045	0.018, 1.529
Sterling et al., 2015	18	0.018	0.071	0.943	-0.471, 0.507
Taylor et al., 2018	44	1.148	5.315	< 0.001	0.725, 1.571
Whitehead, 2005	20	0.898	3.148	0.002	0.339, 1.457
Overall		**0.788**	**5.384**	**< 0.001**	**0.501, 1.074**

Note. All *p* values reflect two-tailed probabilities.

Note. Positive Hedges *g* values indicate parent ratings were higher than self-ratings; negative Hedges values indicate self-ratings were higher than parent ratings.

Table 16.

Study-by-Study Hedges g Values Between Parent-rated and Self-rated Broad Internalizing

Study	***N* pair**	**Hedges *g***	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Bitsika et al., 2015	150	0.032	0.442	0.659	-0.110, 0.173
Foley Nicpon et al., 2010	25	0.513	2.323	0.020	0.080, 0.946
Jamison and Oeth Schuttler, 2015	20	-0.128	-0.557	0.577	-0.580, 0.323
Jepsen et al., 2012	44	0.074	0.537	0.341	-0.095, 0.276
Overall		**0.090**	**0.953**	**0.341**	**-0.095, 0.276**

Note. All *p* values reflect two-tailed probabilities.
Note. Positive Hedges *g* values indicate parent ratings were higher than self-ratings; negative Hedges values indicate self-ratings were higher than parent ratings.

Table 17.

Study-by-Study Hedges g Values Between Teacher-rated and Self-rated Anxiety

Study	***N* pair**	**Hedges *g***	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Barnhill et al., 2000	20	1.352	3.593	< 0.001	0.614, 2.089
Foley Nicpon et al., 2010	25	0.418	1.664	0.096	-0.075, 0.911
Hammond and Hoffman, 2014	7	3.313	2.963	0.003	1.121, 5.505
Jepsen et al., 2012	36	-0.035	-0.165	0.869	-0.450, 0.380
Luxford et al., 2017	18	-0.835	-2.541	0.011	-1.479, -0.191
Ooi et al., 2008	6	-0.629	-1.366	0.172	-1.531, 0.274
Overall		**0.295**	**0.812**	**0.417**	**-0.417, 1.006**

Note. All *p* values reflect two-tailed probabilities.
Note. Positive Hedges *g* values indicate teacher ratings were higher than self-ratings; negative Hedges values indicate self-ratings were higher than teacher ratings.

Table 18.

Study-by-Study Hedges g Values Between Teacher-rated and Self-rated Depression

Study	***N* pair**	**Hedges *g***	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Barnhill et al., 2000	20	0.914	3.079	0.002	0.332, 1.496
Foley Nicpon et al., 2010	25	0.825	3.182	0.001	0.317, 1.334
Hammond and Hoffman, 2014	7	1.683	2.634	0.008	0.431, 2.936
Jepsen et al., 2012	36	-0.334	-1.476	0.140	-0.777, 0.110
Overall		**0.670**	**1.661**	**0.097**	**-0.121, 1.461**

Note. All *p* values reflect two-tailed probabilities.
Note. Positive Hedges *g* values indicate teacher ratings were higher than self-ratings; negative Hedges values indicate self-ratings were higher than teacher ratings.

Table 19.

Study-by-Study Hedges g Values Between Teacher-rated and Self-rated Broad Internalizing

Study	*N* pair	Hedges *g*	Z-Value	*p*-Value	95% Confidence Interval
Foley Nicpon et al., 2010	25	0.343	1.470	0.142	-0.115, 0.801
Jepsen et al., 2012	36	-0.409	-1.751	0.080	-0.867, 0.049
Overall		**-0.033**	**-0.088**	**0.930**	**-0.770, 0.704**

Note. All *p* values reflect two-tailed probabilities.
Note. Positive Hedges *g* values indicate teacher ratings were higher than self-ratings; negative Hedges values indicate self-ratings were higher than teacher ratings.

Table 20.

Study-by-Study Hedges g Values Between Parent-rated and Teacher-rated Anxiety

Study	***N* pair**	**Hedges *g***	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Barnhill et al., 2000	20	-0.043	-0.164	0.870	-0.553, 0.468
Chandler et al., 2016	277	0.615	7.753	< 0.001	0.459, 0.770
Ellison et al., 2015	67	-0.347	-2.300	0.021	0.459, 0.770
Foley Nicpon et al., 2010	33	-0.005	-0.024	0.981	-0.409, 0.399
Hammond and Hoffman, 2014	7	-0.199	-0.494	0.621	-0.991, 0.592
Jepsen et al., 2012	36	0.246	1.227	0.220	-0.147, 0.639
Lane et al., 2013	39	-0.587	-2.831	0.005	-0.993, -0.181
Luxford et al., 2017	18	1.371	3.522	< 0.001	0.608, 2.134
McDonald et al., 2016	118	-0.170	-1.602	0.109	-0.377, 0.038
Ooi et al., 2008	6	-0.825	-1.625	0.104	-1.819, 0.170
Slavin, 2010	6	-0.234	-0.550	0.582	-1.067, 0.599
Overall		**0.156**	**3.120**	**0.002**	**0.058, 0.254**

Note. All *p* values reflect two-tailed probabilities.

Note. Positive Hedges *g* values indicate parent ratings were higher than teacher ratings; negative Hedges values indicate teacher ratings were higher than parent ratings.

Table 21.

Study-by-Study Hedges g Values Between Parent-rated and Teacher-rated Depression

Study	***N* pair**	**Hedges *g***	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Barnhill et al., 2000	20	0.549	1.960	0.050	-0.000, 1.098
Chandler et al., 2016	277	0.964	10.980	< 0.001	0.792, 1.137
Ellison et al., 2015	67	-0.078	-0.530	0.596	-0.364, 0.209
Foley Nicpon et al., 2010	33	0.066	0.321	0.748	-0.337, 0.470
Hammond and Hoffman, 2014	7	0.013	0.032	0.975	-0.767, 0.792
Jepsen et al., 2012	36	0.351	1.830	0.067	-0.025, 0.726
Lane et al., 2013	39	-0.066	-0.402	0.687	-0.389, 0.257
McDonald et al., 2016	118	-0.098	-0.899	0.369	-0.310, 0.115
Slavin, 2010	6	-0.318	-0.740	0.459	-1.162, 0.525
Overall		**0.176**	**0.937**	**0.349**	**-0.192, 0.545**

Note. All *p* values reflect two-tailed probabilities.
Note. Positive Hedges *g* values indicate parent ratings were higher than teacher ratings; negative Hedges values indicate teacher ratings were higher than parent ratings.

Table 22.

Study-by-Study Hedges g Values Between Parent-rated and Teacher-rated Broad Internalizing

Study	***N* pair**	**Hedges *g***	**Z-Value**	***p*-Value**	**95% Confidence Interval**
Barnhill et al., 2000	20	0.329	1.255	0.209	-0.185, 0.843
Chalfant et al., 2007	28	0.656	2.715	0.007	0.182, 1.129
Connolly, 2012	71	0.304	2.047	0.041	0.013, 0.596
Dauterman, 2017	70	0.097	0.913	0.361	-0.111, 0.305
Ellison et al., 2015	67	-0.204	-1.405	0.160	-0.487, 0.080
Foley Nicpon et al., 2010	33	0.135	0.667	0.504	-0.262, 0.532
Jepsen et al., 2012	36	0.483	2.226	0.026	0.058, 0.908
Lane et al., 2013	39	-0.173	-0.925	0.355	-0.540, 0.194
McDonald et al., 2016	118	-0.150	-1.359	0.174	-0.366, 0.066
Peterson, 2017	26	0.119	0.682	0.495	-0.224, 0.462
Rodriguez, 2017	166	0.502	4.613	< 0.001	0.289, 0.715
Rosen et al., 2019	283	0.358	4.933	< 0.001	0.216, 0.500
Slavin, 2010	6	-0.415	-0.961	0.337	-1.262, 0.432
Stratis and Lecavalier, 2017	403	-0.028	-0.471	0.638	-0.143, 0.088
Overall		**0.153**	**2.047**	**0.041**	**0.006, 0.299**

Note. All *p* values reflect two-tailed probabilities.

Note. Positive Hedges *g* values indicate parent ratings were higher than teacher ratings; negative Hedges values indicate teacher ratings were higher than parent ratings.

Table 23.

Information for Studies Included in Analyses for Mean Youth Cognitive Ability as a Moderator for Standardized Mean Differences

Study	**Rater-pair**	**Construct**	***M* FSIQ Score**	**Hedges *g***	***N* pair**
Barnhill et al., 2000	Parent vs. Self	Anxiety	97.94	0.990	20
Bellini, 2004	Parent vs. Self	Anxiety	99.94	0.689	41
Bitsika and Sharpley, 2015	Parent vs. Self	Anxiety	96.00	0.279	139
Bitsika et al., 2019	Parent vs. Self	Anxiety	94.98	-0.125	150
Bitsika et al., 2014	Parent vs. Self	Anxiety	110.40	0.188	32
Boulter et al., 2014	Parent vs. Self	Anxiety	108.50	0.071	170
Chiu et al., 2016	Parent vs. Self	Anxiety	92.10	0.355	28
Chow, 2008	Parent vs. Self	Anxiety	107.66	0.265	32
Clarke et al., 2017	Parent vs. Self	Anxiety	97.71	0.064	14
Foley Nicpon et al., 2010	Parent vs. Self	Anxiety	122.25	0.460	25
Hallett et al., 2013	Parent vs. Self	Anxiety	88.07	-0.050	79
Hollocks et al., 2013	Parent vs. Self	Anxiety	95.60	0.126	38
Jepsen et al., 2012	Parent vs. Self	Anxiety	91.13	0.233	44
Kaat, 2014	Parent vs. Self	Anxiety	90.70	0.196	43
Keith et al., 2018	Parent vs. Self	Anxiety	109.90	-0.366	26
Lopata et al., 2010	Parent vs. Self	Anxiety	110.14	0.579	40
Luxford et al., 2017	Parent vs. Self	Anxiety	105.44	0.410	18
Neil et al., 2019	Parent vs. Self	Anxiety	101.55	-0.086	19

Table 23 (cont'd)

Reaven et al., 2009	Parent vs. Self	Anxiety	102.46	0.472	10
Rosen and Lerner, 2018	Parent vs. Self	Anxiety	102.82	0.580	51
Rump, 2012	Parent vs. Self	Anxiety	106.00	-0.572	19
Sharpley et al., 2015	Parent vs. Self	Anxiety	101.12	0.090	16
Sterling et al., 2015	Parent vs. Self	Anxiety	104.60	-0.190	19
Stern et al., 2014	Parent vs. Self	Anxiety	101.70	0.392	119
Taylor et al., 2018	Parent vs. Self	Anxiety	101.66	0.665	44
Barnhill et al., 2000	Parent vs. Self	Depression	97.94	1.636	20
Chow, 2008	Parent vs. Self	Depression	107.66	0.904	32
Foley Nicpon et al., 2010	Parent vs. Self	Depression	122.25	0.973	25
Hollocks et al., 2013	Parent vs. Self	Depression	95.60	0.983	38
Jepsen et al., 2012	Parent vs. Self	Depression	91.13	0.009	44
Kaat, 2014	Parent vs. Self	Depression	90.70	0.257	41
Lee, 2009	Parent vs. Self	Depression	104.00	0.863	30
Lopata et al., 2010	Parent vs. Self	Depression	110.14	1.152	40
Rosen and Lerner, 2018	Parent vs. Self	Depression	102.82	1.053	51
Rump, 2012	Parent vs. Self	Depression	103.00	0.773	19
Sterling et al., 2015	Parent vs. Self	Depression	104.60	0.018	18
Taylor et al., 2018	Parent vs. Self	Depression	101.66	1.148	44
Bitsika et al., 2019	Parent vs. Self	Internalizing	94.90	0.032	150
Foley Nicpon et al., 2010	Parent vs. Self	Internalizing	122.25	0.513	25

Table 23 (cont'd)

Jepsen et al., 2012	Parent vs. Self	Internalizing	91.12	0.074	44
Barnhill et al., 2000	Parent vs. Teacher	Anxiety	97.94	-0.043	20
Chandler et al., 2016	Parent vs. Teacher	Anxiety	72.70	0.615	277
Ellison et al., 2015	Parent vs. Teacher	Anxiety	83.85	-0.347	67
Foley Nicpon et al., 2010	Parent vs. Teacher	Anxiety	122.25	-0.005	33
Jepsen et al., 2012	Parent vs. Teacher	Anxiety	91.13	0.246	36
Luxford et al., 2017	Parent vs. Teacher	Anxiety	105.44	1.371	18
McDonald et al., 2016	Parent vs. Teacher	Anxiety	103.12	-0.170	118
Barnhill et al., 2000	Parent vs. Teacher	Depression	97.94	0.549	20
Chandler et al., 2016	Parent vs. Teacher	Depression	72.70	0.964	277
Ellison et al., 2015	Parent vs. Teacher	Depression	83.85	-0.078	67
Foley-Nicpon et al., 2010	Parent vs. Teacher	Depression	122.25	0.066	33
Jepsen et al., 2012	Parent vs. Teacher	Depression	91.13	0.351	36
McDonald et al., 2016	Parent vs. Teacher	Depression	103.12	-0.098	118
Barnhill et al., 2000	Parent vs. Teacher	Internalizing	97.94	0.329	20
Ellison et al., 2015	Parent vs. Teacher	Internalizing	83.85	-0.204	67
Foley-Nicpon et al., 2010	Parent vs. Teacher	Internalizing	122.25	0.135	33
Jepsen et al., 2012	Parent vs. Teacher	Internalizing	91.13	0.483	36
McDonald et al., 2016	Parent vs. Teacher	Internalizing	103.12	-0.150	118
Rodriguez, 2017	Parent vs. Teacher	Internalizing	93.86	0.502	166

Table 23 (cont'd)

Rosen et al., 2019	Parent vs. Teacher	Internalizing	85.81	0.358	283
Barnhill et al., 2000	Teacher vs. Self	Anxiety	97.94	1.352	20
Foley-Nicpon et al., 2010	Teacher vs. Self	Anxiety	122.25	0.418	25
Jepsen et al., 2012	Teacher vs. Self	Anxiety	91.13	-0.035	36
Luxford et al., 2017	Teacher vs. Self	Anxiety	105.44	-0.835	18
Barnhill et al., 2000	Teacher vs. Self	Depression	97.94	0.914	20
Foley-Nicpon et al., 2010	Teacher vs. Self	Depression	122.25	0.825	25
Jepsen et al., 2012	Teacher vs. Self	Depression	91.13	-0.334	36
Foley-Nicpon et al., 2010	Teacher vs. Self	Internalizing	122.25	0.343	25
Jepsen et al., 2012	Teacher vs. Self	Internalizing	91.13	-0.409	36

Note. Positive Hedges *g* values indicate parent ratings were higher than teacher ratings; negative Hedges values indicate teacher ratings were higher than parent ratings.

Table 24.

Information for Studies Included in Analyses for Youth Mean Age as a Moderator for Standardized Mean Differences

Study	**Rater-pair**	**Construct**	***M* Age (years)**	**Hedges *g***	***N* pair**
Barnhill et al., 2000	Parent vs. Self	Anxiety	10.70	0.990	20
Bellini, 2004	Parent vs. Self	Anxiety	14.22	0.689	41
Bitsika and Sharpley, 2015	Parent vs. Self	Anxiety	11.20	0.279	139
Bitsika et al., 2019	Parent vs. Self	Anxiety	11.20	-0.125	150
Bitsika et al., 2014	Parent vs. Self	Anxiety	11.20	0.188	32
Blakeley-Smith et al., 2012	Parent vs. Self	Anxiety	10.10	0.382	63
Boulter et al., 2014	Parent vs. Self	Anxiety	12.70	0.071	170
Carruthers et al., 2018	Parent vs. Self	Anxiety	12.88	0.137	38
Chalfant et al., 2007	Parent vs. Self	Anxiety	10.80	0.217	28
Chiu et al., 2016	Parent vs. Self	Anxiety	12.00	0.355	28
Chow, 2008	Parent vs. Self	Anxiety	10.28	0.265	32
Clarke et al., 2017	Parent vs. Self	Anxiety	12.64	0.064	14
Conaughton et al., 2017	Parent vs. Self	Anxiety	9.74	-0.667	21
Freeman, 2009	Parent vs. Self	Anxiety	11.96	0.314	61
Hallett et al., 2013	Parent vs. Self	Anxiety	13.50	-0.050	79
Hammond and Hoffman, 2014	Parent vs. Self	Anxiety	13.90	2.585	10
Hollocks et al., 2013	Parent vs. Self	Anxiety	12.90	0.126	38
Jepsen et al., 2012	Parent vs. Self	Anxiety	15.06	0.233	44

Table 24 (cont'd)

Joyce et al., 2017	Parent vs. Self	Anxiety	16.81	0.001	13
Kaat, 2014	Parent vs. Self	Anxiety	12.40	0.196	43
Keith et al., 2018	Parent vs. Self	Anxiety	14.20	-0.366	26
Lopata et al., 2010	Parent vs. Self	Anxiety	9.75	0.579	40
Luxford et al., 2017	Parent vs. Self	Anxiety	13.20	0.410	18
Magiati et al., 2014	Parent vs. Self	Anxiety	12.10	-0.426	38
Mertens et al., 2017	Parent vs. Self	Anxiety	13.80	0.083	22
Neil et al., 2019	Parent vs. Self	Anxiety	10.24	-0.086	19
Ooi et al., 2008	Parent vs. Self	Anxiety	11.50	-0.993	6
Ooi et al., 2016	Parent vs. Self	Anxiety	11.21	-0.608	70
Reaven et al., 2009	Parent vs. Self	Anxiety	11.10	0.472	10
Rodgers et al., 2016	Parent vs. Self	Anxiety	11.10	-0.078	157
Rosen and Lerner, 2018	Parent vs. Self	Anxiety	12.15	0.580	51
Rump, 2012	Parent vs. Self	Anxiety	14.68	-0.572	19
Sharpley et al., 2015	Parent vs. Self	Anxiety	11.43	0.090	16
Sterling et al., 2015	Parent vs. Self	Anxiety	14.45	-0.190	19
Stern et al., 2014	Parent vs. Self	Anxiety	12.30	0.392	119
Storch, 2015	Parent vs. Self	Anxiety	12.75	1.809	16
Taylor et al., 2018	Parent vs. Self	Anxiety	10.33	0.665	44
Van Schalkwyk et al., 2018	Parent vs. Self	Anxiety	16.40	-0.091	35
Whitehead, 2005	Parent vs. Self	Anxiety	14.85	1.537	20
Wijnhoven et al., 2018	Parent vs. Self	Anxiety	11.25	0.002	168

Table 24 (cont'd)

Wood et al., 2015	Parent vs. Self	Anxiety	9.18	2.220	14
Wood et al., 2009	Parent vs. Self	Anxiety	12.40	0.810	14
Barnhill et al., 2000	Parent vs. Self	Depression	10.70	1.636	20
Bitsika et al., 2019	Parent vs. Self	Depression	11.20	-0.189	150
Chow, 2008	Parent vs. Self	Depression	10.28	0.904	32
Freeman, 2009	Parent vs. Self	Depression	11.96	0.963	61
Hammond and Hoffman, 2014	Parent vs. Self	Depression	13.90	1.738	12
Hollocks et al., 2013	Parent vs. Self	Depression	12.90	0.983	38
Jepsen et al., 2012	Parent vs. Self	Depression	15.06	0.009	44
Kaat, 2014	Parent vs. Self	Depression	12.40	0.257	41
Lee, 2009	Parent vs. Self	Depression	9.33	0.863	30
Lopata et al., 2010	Parent vs. Self	Depression	9.75	1.152	40
Richdale and Baglin, 2015	Parent vs. Self	Depression	10.03	1.001	17
Rosen and Lerner, 2018	Parent vs. Self	Depression	12.15	1.053	51
Rump, 2012	Parent vs. Self	Depression	14.68	0.773	19
Sterling et al., 2015	Parent vs. Self	Depression	14.45	0.018	18
Taylor et al., 2018	Parent vs. Self	Depression	10.33	1.148	44
Whitehead, 2005	Parent vs. Self	Depression	14.85	0.898	20
Bitsika et al., 2019	Parent vs. Self	Internalizing	11.2	0.032	150
Jamison and Oeth Schuttler, 2015	Parent vs. Self	Internalizing	16.04	-0.128	20
Jepsen et al., 2012	Parent vs. Self	Internalizing	15.06	0.074	44

Table 24 (cont'd)

Barnhill et al., 2000	Parent vs. Teacher	Anxiety	10.70	-0.043	20
Chandler et al., 2016	Parent vs. Teacher	Anxiety	6.00	0.615	277
Ellison et al., 2015	Parent vs. Teacher	Anxiety	8.23	-0.347	67
Hammond and Hoffman, 2014	Parent vs. Teacher	Anxiety	13.90	-0.199	7
Jepsen et al., 2012	Parent vs. Teacher	Anxiety	15.06	0.246	36
Lane et al., 2013	Parent vs. Teacher	Anxiety	4.30	-0.587	39
Luxford et al., 2017	Parent vs. Teacher	Anxiety	13.20	1.371	18
McDonald et al., 2016	Parent vs. Teacher	Anxiety	8.74	-0.170	118
Ooi et al., 2008	Parent vs. Teacher	Anxiety	11.50	-0.825	6
Barnhill et al., 2000	Parent vs. Teacher	Depression	10.70	0.549	20
Chandler et al., 2016	Parent vs. Teacher	Depression	6.00	0.964	277
Ellison et al., 2015	Parent vs. Teacher	Depression	8.23	-0.078	67
Hammond and Hoffman, 2014	Parent vs. Teacher	Depression	13.90	0.013	7
Jepsen et al., 2012	Parent vs. Teacher	Depression	15.06	0.351	36
Lane et al., 2013	Parent vs. Teacher	Depression	4.30	-0.066	39
McDonald et al., 2016	Parent vs. Teacher	Depression	8.74	-0.098	118
Barnhill et al., 2000	Parent vs. Teacher	Internalizing	10.7	0.329	20
Chalfant et al., 2007	Parent vs. Teacher	Internalizing	10.8	0.656	28
Connolly, 2012	Parent vs. Teacher	Internalizing	9.3	0.304	71
Dauterman, 2017	Parent vs. Teacher	Internalizing	4.8	0.097	70

Table 24 (cont'd)

Ellison et al., 2015	Parent vs. Teacher	Internalizing	8.2	-0.204	67
Jepsen et al., 2012	Parent vs. Teacher	Internalizing	15.1	0.483	36
Lane et al., 2013	Parent vs. Teacher	Internalizing	4.3	-0.173	39
McDonald et al., 2016	Parent vs. Teacher	Internalizing	8.7	-0.150	118
Peterson, 2017	Parent vs. Teacher	Internalizing	4.9	0.119	26
Rodriguez, 2017	Parent vs. Teacher	Internalizing	5.1	0.502	166
Rosen et al., 2019	Parent vs. Teacher	Internalizing	10.5	0.358	283
Stratis and Lecavalier, 2017	Parent vs. Teacher	Internalizing	10.5	-0.028	403
Barnhill et al., 2000	Teacher vs. Self	Anxiety	10.70	1.352	20
Hammond and Hoffman, 2014	Teacher vs. Self	Anxiety	13.90	3.313	7
Jepsen et al., 2012	Teacher vs. Self	Anxiety	15.06	-0.035	36
Luxford et al., 2017	Teacher vs. Self	Anxiety	13.20	-0.835	18
Ooi et al., 2008	Teacher vs. Self	Anxiety	11.50	-0.629	6
Barnhill et al., 2000	Teacher vs. Self	Depression	10.70	0.914	20
Hammond and Hoffman, 2014	Teacher vs. Self	Depression	13.90	1.683	7
Jepsen et al., 2012	Teacher vs. Self	Depression	15.06	-0.334	36
Jepsen et al., 2012	Teacher vs. Self	Internalizing	15.06	-0.409	36

Note. Positive Hedges *g* values indicate parent ratings were higher than self-ratings; negative Hedges values indicate self-ratings were higher than parent ratings. Positive Hedges *g* values indicate teacher ratings were higher than self-ratings; negative Hedges values indicate self-ratings were higher than teacher ratings. Positive Hedges *g* values indicate parent ratings were higher than teacher ratings; negative Hedges values indicate teacher ratings were higher than parent ratings.

Table 25.

Information for Studies Included in Analyses for Method of Self-Report Administration as a Moderator for Standardized Mean Differences

Study	Rater-pair	Construct	Method of Self-Report Administration	Hedges *g*	*N* pair
Bitsika and Sharpley, 2015	Parent vs. Self	Anxiety	Assessment completed at home	0.279	139
Bitsika et al., 2019	Parent vs. Self	Anxiety	Assessment completed at home	-0.125	150
Chalfant et al., 2007	Parent vs. Self	Anxiety	Assessment completed at home	0.217	28
Elzinga, 2015	Parent vs. Self	Anxiety	Assessment completed at home	1.149	26
Freeman, 2009	Parent vs. Self	Anxiety	Assessment completed at home	0.314	61
Hallett et al., 2013	Parent vs. Self	Anxiety	Assessment completed at home	-0.050	79
Jepsen et al., 2012	Parent vs. Self	Anxiety	Assessment completed at home	0.233	44
Magiati et al., 2014	Parent vs. Self	Anxiety	Assessment completed at home	-0.426	38
Rodgers et al., 2016	Parent vs. Self	Anxiety	Assessment completed at home	-0.078	157
Barnhill et al., 2000	Parent vs. Self	Anxiety	Assessment completed in clinic	0.990	20
Foley Nicpon et al., 2010	Parent vs. Self	Anxiety	Assessment completed in clinic	0.460	25
Hammond and Hoffman, 2014	Parent vs. Self	Anxiety	Assessment completed in clinic	2.585	10
Hollocks et al., 2013	Parent vs. Self	Anxiety	Assessment completed in clinic	0.126	38
Mertens et al., 2017	Parent vs. Self	Anxiety	Assessment completed in clinic	0.083	22
Ooi et al., 2016	Parent vs. Self	Anxiety	Assessment completed in clinic	-0.608	70
Sterling et al., 2015	Parent vs. Self	Anxiety	Assessment completed in clinic	-0.190	19
Wood et al., 2015	Parent vs. Self	Anxiety	Assessment completed in clinic	2.220	14
Bellini, 2004	Parent vs. Self	Anxiety	Assessment read to child in clinic	0.689	41

Table 25 (cont'd)

Blakeley-Smith et al., 2012	Parent vs. Self	Anxiety	Assessment read to child in clinic	0.382	63
Chow, 2008	Parent vs. Self	Anxiety	Assessment read to child in clinic	0.265	32
Lopata et al., 2010	Parent vs. Self	Anxiety	Assessment read to child in clinic	0.579	40
Bitsika et al., 2019	Parent vs. Self	Depression	Assessment completed at home	-0.189	150
Freeman, 2009	Parent vs. Self	Depression	Assessment completed at home	0.963	61
Jepsen et al., 2012	Parent vs. Self	Depression	Assessment completed at home	0.009	44
Richdale and Baglin, 2015	Parent vs. Self	Depression	Assessment read to child at home	1.001	17
Barnhill et al., 2000	Parent vs. Self	Depression	Assessment completed in clinic	1.636	20
Foley Nicpon et al., 2010	Parent vs. Self	Depression	Assessment completed in clinic	0.973	25
Hammond and Hoffman, 2014	Parent vs. Self	Depression	Assessment completed in clinic	1.738	12
Hollocks et al., 2013	Parent vs. Self	Depression	Assessment completed in clinic	0.983	38
Sterling et al., 2015	Parent vs. Self	Depression	Assessment completed in clinic	0.018	18
Chow, 2008	Parent vs. Self	Depression	Assessment read to child in clinic	0.904	32
Lee, 2009	Parent vs. Self	Depression	Assessment read to child in clinic	0.863	30
Lopata et al., 2010	Parent vs. Self	Depression	Assessment read to child in clinic	1.152	40
Bitsika et al., 2019	Parent vs. Self	Internalizing	Assessment completed at home	0.032	150
Jepsen et al., 2012	Parent vs. Self	Internalizing	Assessment completed at home	0.074	44
Foley Nicpon et al., 2010	Parent vs. Self	Internalizing	Assessment completed in clinic	0.513	25

Table 26.

Information for Studies Included in Analyses for Correlation Coefficient as a Moderator for Correlations

Study	Rater-pair	Construct	*N* pair	Correlation Coefficient	Correlation Value
Bermudez et al., 2015	Parent vs. Self	Anxiety	38	Pearson's *r*	0.319
Bitsika et al., 2019	Parent vs. Self	Anxiety	150	Pearson's *r*	0.570
Chow, 2008	Parent vs. Self	Anxiety	32	Pearson's *r*	-0.022
Farrugia and Hudson, 2006	Parent vs. Self	Anxiety	29	Pearson's *r*	0.697
Freeman, 2009	Parent vs. Self	Anxiety	61	Pearson's *r*	0.550
Hallett et al., 2013	Parent vs. Self	Anxiety	79	Pearson's *r*	0.490
Jepsen et al., 2012	Parent vs. Self	Anxiety	40	Pearson's *r*	0.134
Lohr et al., 2017	Parent vs. Self	Anxiety	41	Pearson's *r*	0.230
Lopata et al., 2010	Parent vs. Self	Anxiety	73	Pearson's *r*	0.430
Pisula et al., 2017	Parent vs. Self	Anxiety	35	Pearson's *r*	0.420
Rosen and Lerner, 2018	Parent vs. Self	Anxiety	51	Pearson's *r*	0.050
Rosenberg, 2016	Parent vs. Self	Anxiety	20	Pearson's *r*	0.590
Rump, 2012	Parent vs. Self	Anxiety	19	Pearson's *r*	0.190
Schiltz et al., 2018	Parent vs. Self	Anxiety	53	Pearson's *r*	0.370
Schwartz, 2010	Parent vs. Self	Anxiety	30	Pearson's *r*	0.210
Sterling et al., 2015	Parent vs. Self	Anxiety	67	Pearson's *r*	0.260
Taylor et al., 2018	Parent vs. Self	Anxiety	44	Pearson's *r*	0.198
Whitehead, 2005	Parent vs. Self	Anxiety	20	Pearson's *r*	0.480
Average *r*					**0.343**
Hurtig et al., 2009	Parent vs. Self	Anxiety	45	Spearman's *rho*	0.290
Mertens et al., 2017	Parent vs. Self	Anxiety	22	Spearman's *rho*	0.270
Average *r*					**0.280**

Table 26 (cont'd)

Blakeley-Smith et al., 2012	Parent vs. Self	Anxiety	63	ICC	0.520
Kaat, 2014	Parent vs. Self	Anxiety	44	ICC	0.529
Magiati, 2014	Parent vs. Self	Anxiety	38	ICC	0.690
Ooi et al., 2016	Parent vs. Self	Anxiety	70	ICC	0.380
Ozsivadjian et al., 2014	Parent vs. Self	Anxiety	30	ICC	0.590
Average *r*					**0.542**
Bohnert et al., 2016	Parent vs. Self	Depression	127	Pearson's *r*	0.610
Chow, 2008	Parent vs. Self	Depression	32	Pearson's *r*	0.292
Freeman, 2009	Parent vs. Self	Depression	61	Pearson's *r*	0.480
Hammond and Hammond, 2014	Parent vs. Self	Depression	12	Pearson's *r*	0.610
Kaat, 2014	Parent vs. Self	Depression	41	Pearson's *r*	0.360
Lee, 2009	Parent vs. Self	Depression	30	Pearson's *r*	0.280
Lopata et al., 2010	Parent vs. Self	Depression	40	Pearson's *r*	0.315
Pisula et al., 2017	Parent vs. Self	Depression	35	Pearson's *r*	0.420
Rosen and Lerner, 2018	Parent vs. Self	Depression	51	Pearson's *r*	0.140
Rosenberg, 2016	Parent vs. Self	Depression	20	Pearson's *r*	0.450
Rump, 2012	Parent vs. Self	Depression	19	Pearson's *r*	-0.160
Taylor et al., 2018	Parent vs. Self	Depression	44	Pearson's *r*	0.368
Vickerstaff et al., 2007	Parent vs. Self	Depression	22	Pearson's *r*	0.670
Average *r*					**0.372**
Hurtig et al., 2009	Parent vs. Self	Depression	45	Spearman's *rho*	0.290
Average *r*					**0.290**
Jepsen et al., 2012	Parent vs. Self	Depression	44	ICC	0.524
Ozsivadjian et al., 2014	Parent vs. Self	Depression	30	ICC	0.620
Average *r*					**0.572**
Bitsika et al., 2019	Parent vs. Self	Internalizing	150	Pearson's *r*	0.606

Table 26 (cont'd)

Pisula et al., 2017	Parent vs. Self	Internalizing	35	Pearson's *r*	0.310
Average *r*					**0.458**
Hurtig et al., 2009	Parent vs. Self	Internalizing	45	Spearman's *rho*	0.270
Average *r*					**0.270**
Jepsen et al., 2012	Parent vs. Self	Internalizing	44	ICC	0.564
Kaat and Lecavalier, 2015	Parent vs. Self	Internalizing	46	ICC	0.250
Average *r*					**0.407**
Kanne et al., 2009	Parent vs. Teacher	Anxiety	177	Pearson's *r*	0.140
Lane et al., 2013	Parent vs. Teacher	Anxiety	39	Pearson's *r*	0.260
McDonald et al., 2016	Parent vs. Teacher	Anxiety	118	Pearson's *r*	0.340
Average *r*					**0.247**
Adams et al., 2018	Parent vs. Teacher	Anxiety	92	Spearman's *rho*	0.390
Hurtig et al., 2009	Parent vs. Teacher	Anxiety	22	Spearman's *rho*	0.190
Average *r*					**0.290**
Jepsen et al., 2012	Parent vs. Teacher	Anxiety	36	ICC	0.286
Ung et al., 2017	Parent vs. Teacher	Anxiety	32	ICC	0.410
Average *r*					**0.348**
Kanne et al., 2009	Parent vs. Teacher	Depression	177	Pearson's *r*	0.080
Lane et al., 2013	Parent vs. Teacher	Depression	39	Pearson's *r*	0.450
McDonald et al., 2016	Parent vs. Teacher	Depression	118	Pearson's *r*	0.300
Vickerstaff et al., 2007	Parent vs. Teacher	Depression	22	Pearson's *r*	0.220
Average *r*					**0.263**
Hurtig et al., 2009	Parent vs. Teacher	Depression	22	Spearman's *rho*	0.310

Table 26 (cont'd)

Average *r*					**0.310**
Jepsen et al., 2012	Parent vs. Teacher	Depression	36	ICC	0.351
Ung et al., 2017	Parent vs. Teacher	Depression	32	ICC	0.330
Average *r*					**0.341**
Connolly, 2012	Parent vs. Teacher	Internalizing	71	Pearson's *r*	0.235
Dauterman, 2017	Parent vs. Teacher	Internalizing	70	Pearson's *r*	0.600
Lane et al., 2013	Parent vs. Teacher	Internalizing	39	Pearson's *r*	0.300
McDonald et al., 2016	Parent vs. Teacher	Internalizing	118	Pearson's *r*	0.280
Peterson, 2017	Parent vs. Teacher	Internalizing	26	Pearson's *r*	0.580
Rodriguez, 2017	Parent vs. Teacher	Internalizing	166	Pearson's *r*	0.120
Average *r*					**0.353**
Hurtig et al., 2009	Parent vs. Teacher	Internalizing	22	Spearman's *rho*	0.050
Average *r*					**0.050**
Jepsen et al., 2012	Parent vs. Teacher	Internalizing	36	ICC	0.212
Ung et al., 2017	Parent vs. Teacher	Internalizing	32	ICC	0.180
Average *r*					**0.196**
Hurtig et al., 2009	Teacher vs. Self	Anxiety	23	Spearman's *rho*	0.340
Average *r*					**0.340**
Jepsen et al., 2012	Teacher vs. Self	Anxiety	36	ICC	0.158
Average *r*					**0.158**
Vickerstaff et al., 2007	Teacher vs. Self	Depression	22	Pearson's *r*	0.220
Average *r*					**0.220**
Hurtig et al., 2009	Teacher vs. Self	Depression	23	Spearman's *rho*	0.660
Average *r*					**0.660**
Jepsen et al., 2012	Teacher vs. Self	Depression	36	ICC	0.091

Table 26 (cont'd)

Average *r*					**0.091**
Hurtig et al., 2009	Teacher vs. Self	Internalizing	23	Spearman's *rho*	0.560
Average *r*					**0.560**
Jepsen et al., 2012	Teacher vs. Self	Internalizing	36	ICC	0.056
Average *r*					**0.056**

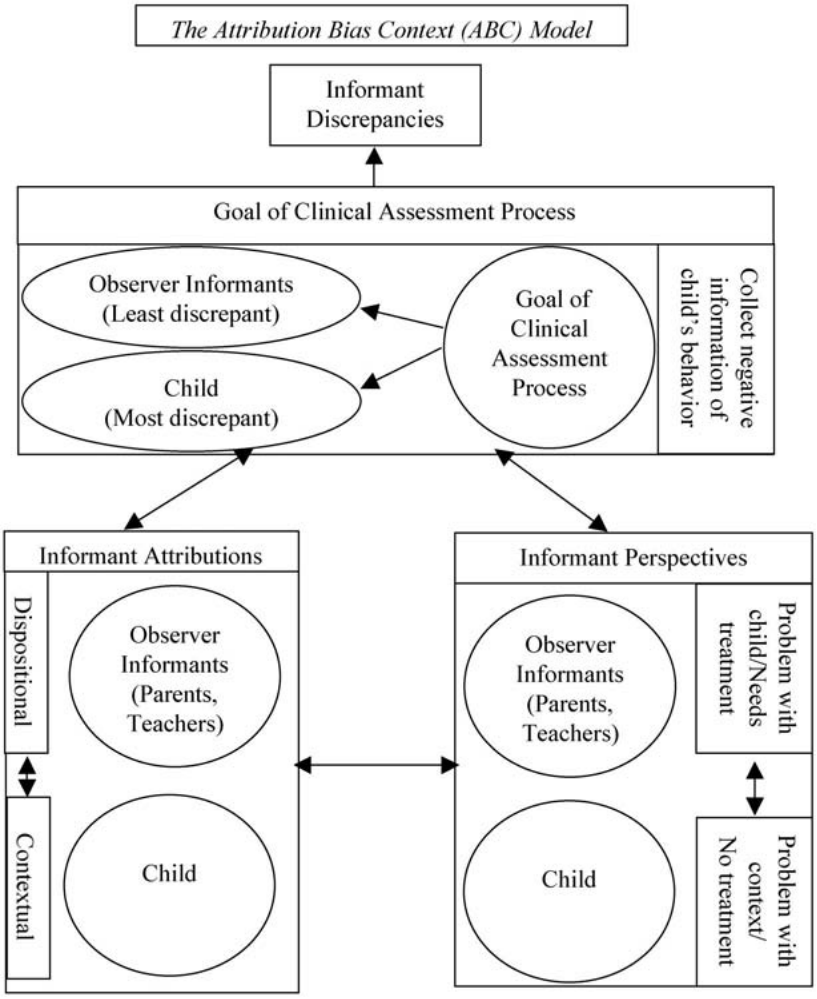


Figure 1. Conceptualization of the Attributional Bias Context (ABC) Model (De Los Reyes & Kazdin, 2005)

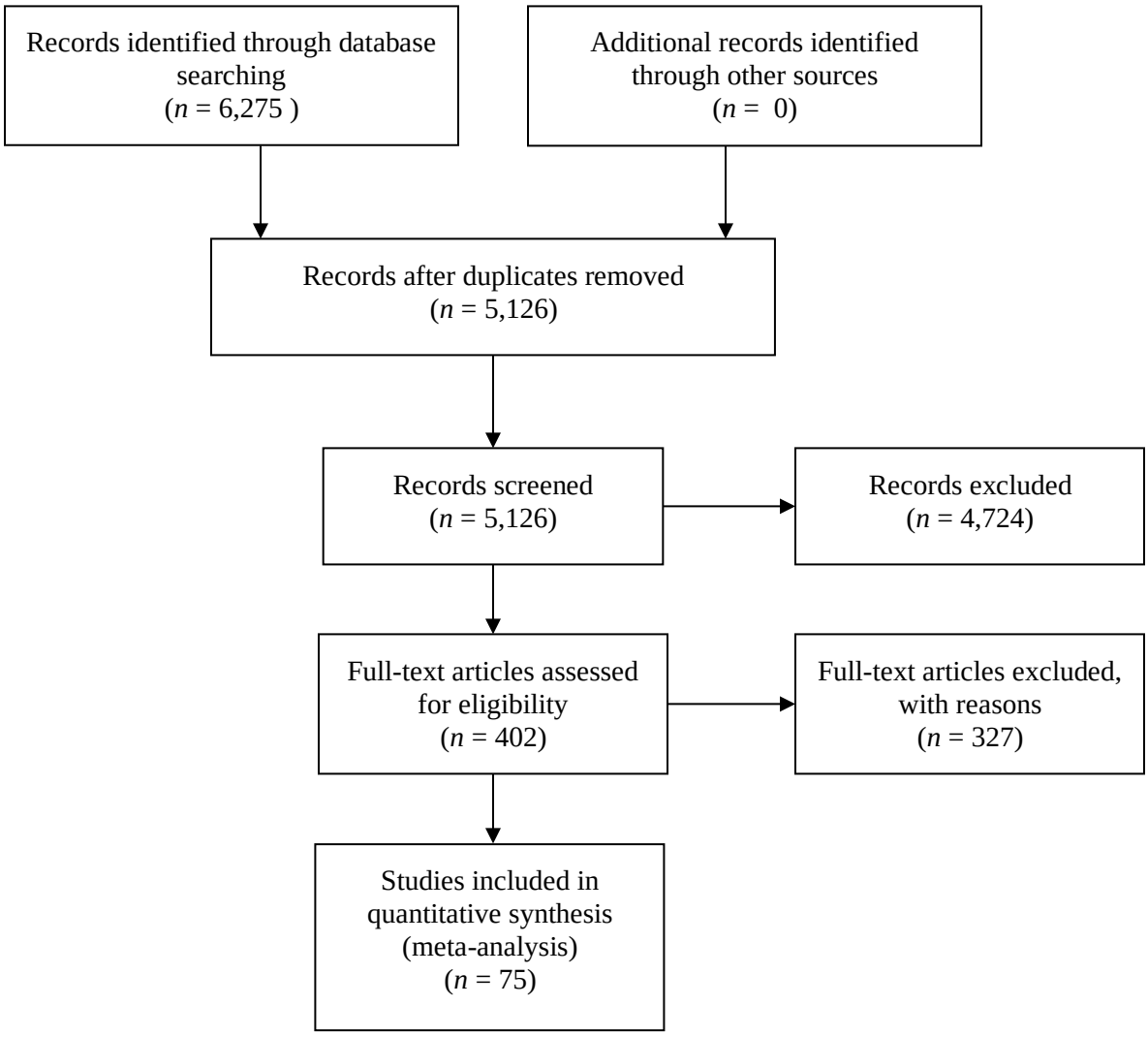


Figure 2. PRISMA (Moher et al., 2009) Flow Diagram for Present Meta-Analysis

APPENDICES

Appendix A

Table 27.

Abstract Screening Criteria Checklist

Criteria	Yes/No Checklist
1. Study is published in English	
2. Study utilizes an ASD sample	
3. Study uses and ASD youth sample (i.e., age 18 years or younger)	
4. Study measures anxiety, depression, or internalizing problems	
5. Study includes youth self-report and parent report, youth self-report and teacher report, or parent ratings and teacher report (study includes data from at least one rater-pair)	

Appendix B

Table 28.

Full-text Screening Criteria Checklist

Criteria	Yes/No Checklist
1. The study includes one or more rating scales that assess depression, anxiety, or internalizing problems.	
2. The study includes multiple informants, specifically at least one rater-pair. Target rater-pairs include youth-parent, youth-teacher, or parent-teacher.	
3. Either (a) for each measure and each rater, the sample size, mean, and standard deviation were reported or (b) critical information about the correlation (association/agreement) between rater-pairs' scores is available (i.e., sample size for the rater-pair and correlation coefficient).	

Appendix C

Table 29.

Coding Sheet

General Article Information	
1. ID Number	
2. Article Name	
3. Study Authors	
4. Publication Year	
5. State/ Country study was conducted	
6. Type of Study	1 = Journal article 2 = Book or book chapter 3 = Dissertation 4 = Master's Thesis 5 = Unpublished Report 888 = Other (*please specify*):
7. Name of Source	Journal name, book name, or name of university where dissertation was produced
8. Study Design	1 = One group observational study 2 = Multiple group observational study 3 = Large data based study 4 = Experimental Research 888 = Other (*please specify):* 999 = Not reported (NR) *circle all that apply
9. Type of Informant/ Rater-pairs	1 = Self-report/ Parent report 2 = Parent report/ Teacher report 3 = Self-report/ Teacher report 888 = Other (*please specify*):

Table 29 (cont'd)

10. Diagnosis in ASD Group	1 = ASD 2 = HFASD 3 = HFA 4 = PDD-NOS 5 = Asperger Syndrome 6 = Autistic disorder 888 = Other (*please specify*): 999 = NR
11. How was the diagnosis made?	1 = DSM-III 2 = DSM-III-R 3 = DSM- IV 4 = DSM-IV-TR 5 = DSM- 5 6 = ICD-9 7 = ICD-10 8 = ADOS/ ADOS-2 involved 9 = ADIR involved 888 = Other (*please specify*): 999= NR * circle all that apply

Group Specific Information	**ASD Group**	**Comparison Group**
Age Range	Minimum age: Maximum age:	Minimum age: Maximum age:
Age Mean	In years:	In years:
Age *SD*		
Ethnicity	% White/non-Hispanic: % Black/African American: % Hispanic or Latino: % Asian/Asian American: % American Indian/Alaska Native: % Native Hawaiian or Pacific Islander: % Other (*please specify*):	% White/non-Hispanic: % Black/African American: % Hispanic or Latino: % Asian/Asian American: % American Indian/Alaska Native: % Native Hawaiian or Pacific Islander: % Other (*please specify*):

Table 29 (cont'd)

Gender (%)	% Male = % Female =	% Male = % Female =
Participant IQ	Measure = Mean = Minimum = Maximum = *SD* = 999 = NR	Measure = Mean = Minimum = Maximum = *SD* = 999 = NR
Participant adaptive functioning	Measure = Mean = Minimum = Maximum = *SD* = 999 = NR	Mean = Minimum = Maximum = *SD* = 999 = NR
Participant language functioning	Measure = Mean = Minimum = Maximum = *SD* = 999 = NR	Mean = Minimum = Maximum = *SD* = 999 = NR
SES (central tendency [mean if possible], *SD)*	Parent education = Free/ Subsidized = Income-based = Other (*please specify*): 999= NR	Parent education = Free/ Subsidized = Income-based = Other (*please specify*): 999= NR
Social desirability	Measure = Total score = 999 = NR	Measure = Total score = 999 = NR
Parent depression	Measure = Depression score = 999 = NR	Measure = Depression score = 999 = NR
Parent stress	Measure = Stress score = 999 = NR	Measure = Stress score = 999 = NR

Table 29 (cont'd)

Measure(s) used for ***parent report***	**Measure # 1** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 2** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 3** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 4** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info =	**Measure # 1** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 2** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 3** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 4** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info =

Table 29 (cont'd)

	Measure # 5 Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info =	**Measure # 5** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info =
Measure(s) used for ***self-report***	**Measure # 1** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 2** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 3** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 4** Measure = Construct = Score (total, subscale, etc.) = *N* =	**Measure # 1** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 2** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 3** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 4** Measure = Construct = Score (total, subscale, etc.) = *N* =

Table 29 (cont'd)

	Mean = *SD* = Descriptive Category = Other notable info = **Measure # 5** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info =	Mean = *SD* = Descriptive Category = Other notable info = **Measure # 5** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info =
Location self-report completed	1 = Clinic/school 2 = Home 3 = Research setting 888 = Other (*please specify*) 999 = NR	1 = Clinic/school 2 = Home 3 = Research setting 888 = Other (please specify) 999 = NR
Method of self-report completion	1 = Assessment read to child in clinic 2 = Assessment read to child at home 3 = Assessment completed in clinic 4 = Assessment completed at home 888 = Other (*please specify*) 999 = NR	1 = Youth completes unassisted 2 = Clinician/researcher reads items to all youth in sample 3 = Clinician/research reads items to youth as needed 4 = Parent read items to all youth in sample 5 = Parent read items to youth as needed/at their discretion 888 = Other (*please specify*) 999 = NR
Youth rating scale training completed	1 = yes 2 = no	

Table 29 (cont'd)

Measure(s) used for ***teacher report***	**Measure # 1** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 2** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 3** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 4** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 5** Measure = Construct = Score (total, subscale, etc.) =	**Measure # 1** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 2** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 3** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 4** Measure = Construct = Score (total, subscale, etc.) = *N* = Mean = *SD* = Descriptive Category = Other notable info = **Measure # 5** Measure = Construct = Score (total, subscale, etc.) =

Table 29 (cont'd)

	N = Mean = *SD* = Descriptive Category = Other notable info =	*N* = Mean = *SD* = Descriptive Category = Other notable info =

ASD Group

Correlations	**Self vs. parent**	**Parent vs. teacher**	**Teacher vs. self**
Measure # 1 = Construct =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =
Measure # 2 = Construct =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =
Measure # 3 = Construct =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =
Measure # 4 = Construct =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pair = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pair = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =
Measure # 5 = Construct =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =

Table 29 (cont'd)

	Comparison Group		
Correlations	**Self vs. parent**	**Parent vs. teacher**	**Teacher vs. self**
Measure # 1 = Construct =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =
Measure # 2 = Construct =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =
Measure # 3 = Construct =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =
Measure # 4 = Construct =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =
Measure # 5 = Construct =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =	*N* pairs = Pearson's *r* = Other (specify*)*: Mean difference = *SD* of paired difference =

Other Comments:

Appendix D

Table 30.

Study-by-Study Information for Correlation Between Ratings of Anxiety, Depression, and Broad Internalizing

Study	**Rater-pair**	**Construct**	**Measure(s)**	***N* pair**	**Correlation Coefficient**
Adams et al., 2018	Parent vs. Teacher	Anxiety	The Anxiety Scale for Children with Autism Spectrum Disorder	92	Spearman's *rho*
Bermudez et al., 2015	Parent vs. Self	Anxiety	SCARED	38	Pearson's *r*
Bitsika et al., 2019	Parent vs. Self	Anxiety	CASI-4	150	Pearson's *r*
		Internalizing	CASI-4	150	Pearson's *r*
Blakeley-Smith et al., 2012	Parent vs. Self	Anxiety	SCARED	63	ICC
Bohnert et al., 2016	Parent vs. Self	Depression	ASEBA CBCL and YSR	127	Pearson's *r*
Chow, 2008	Parent vs. Self	Anxiety	BASC-2 PRS and SRP	32	Pearson's *r*
		Depression	BASC-2 PRS and CDI	32	Pearson's *r*
Connolly, 2012	Parent vs. Teacher	Internalizing	ASEBA CBCL and TRF	71	Pearson's *r*
Dauterman, 2017	Parent vs. Teacher	Internalizing	BASC-2 PRS and TRS	70	Pearson's *r*
Farrugia and Hudson, 2006	Parent vs. Self	Anxiety	SCAS	29	Pearson's *r*
Freeman, 2009	Parent vs. Self	Anxiety	Revised Children's Manifest Anxiety Scale (RCMAS) vs. BASC-2	61	Pearson's *r*
		Depression	BASC-2 PRS and CDI	61	Pearson's *r*
Hallett et al., 2013	Parent vs. Self	Anxiety	RCADS	79	Pearson's *r*
Hurtig et al., 2009	Parent vs. Self	Anxiety	ASEBA CBCL and YSR	45	Spearman's *rho*
		Depression	ASEBA CBCL and YSR	45	Spearman's *rho*
		Internalizing	ASEBA CBCL and YSR	45	Spearman's *rho*

Table 30 (cont'd)

	Parent vs. Teacher	Anxiety	ASEBA CBCL and TRF	22	Spearman's *rho*
		Depression	ASEBA CBCL and TRF	22	Spearman's *rho*
		Internalizing	ASEBA CBCL and TRF	22	Spearman's *rho*
	Teacher vs. Self	Anxiety	ASEBA CBCL and YSR	23	Spearman's *rho*
		Depression	ASEBA CBCL and YSR	23	Spearman's *rho*
		Internalizing	ASEBA CBCL and YSR	23	Spearman's *rho*
Jepsen et al., 2012	Parent vs. Self	Anxiety	ASEBA CBCL and YSR	44	ICC
		Depression	ASEBA CBCL and YSR	44	ICC
		Internalizing	ASEBA CBCL and YSR	44	ICC
	Parent vs. Teacher	Anxiety	ASEBA CBCL and TRF	36	ICC
		Depression	ASEBA CBCL and TRF	36	ICC
		Internalizing	ASEBA CBCL and TRF	36	ICC
	Teacher vs. Self	Anxiety	ASEBA CBCL and YSR	36	ICC
		Depression	ASEBA CBCL and YSR	36	ICC
		Internalizing	ASEBA CBCL and YSR	36	ICC
Kaat, 2014	Parent vs. Self	Anxiety	RCADS	41	Pearson's *r*
		Depression	RCADS	41	Pearson's *r*
Kaat and Lecavalier, 2015	Parent vs. Self	Internalizing	RCADS	46	ICC
Kanne et al., 2009	Parent vs. Teacher	Anxiety	ASEBA CBCL and TRF	177	Pearson's *r*
		Depression	ASEBA CBCL and TRF	177	Pearson's *r*
Lane et al., 2013	Parent vs. Teacher	Anxiety	ASEBA CBCL and TRF	39	Pearson's *r*

Table 30 (cont'd)

		Depression	ASEBA CBCL and TRF	39	Pearson's *r*
		Internalizing	ASEBA CBCL and TRF	39	Pearson's *r*
Lee, 2009	Parent vs. Self	Depression	BASC-2 PRS and SRP	30	Pearson's *r*
Lohr et al., 2017	Parent vs. Self	Anxiety	SCARED	73	Pearson's *r*
Lopata et al., 2010	Parent vs. Self	Anxiety	BASC-2 PRS and SRP	40	Pearson's *r*
		Depression	BASC-2 PRS and SRP	40	Pearson's *r*
Magiati et al., 2014	Parent vs. Self	Anxiety	SCAS	38	ICC
McDonald et al., 2016	Parent vs. Teacher	Anxiety	BASC-2 PRS and TRS	118	Pearson's *r*
		Depression	BASC-2 PRS and TRS	118	Pearson's *r*
		Internalizing	BASC-2 PRS and TRS	118	Pearson's *r*
Mertens et al., 2017	Parent vs. Self	Anxiety	SCARED	22	Spearman's *rho*
Ooi et al., 2016	Parent vs. Self	Anxiety	SCAS	70	ICC
Ozsivadjian et al., 2014	Parent vs. Self	Anxiety	SCAS	30	ICC
		Depression	CDI	30	ICC
Peterson, 2017	Parent vs. Teacher	Internalizing	BASC-2 PRS and TRS	26	Pearson's *r*
Pisula et al., 2017	Parent vs. Self	Anxiety	ASEBA CBCL and YSR	35	Pearson's *r*
		Depression	ASEBA CBCL and YSR	35	Pearson's *r*
		Internalizing	ASEBA CBCL and YSR	35	Pearson's *r*
Rodriguez, 2017	Parent vs. Teacher	Internalizing	ASEBA CBCL and TRF	166	Pearson's *r*
Rosen and Lerner, 2018	Parent vs. Self	Anxiety	MASC	51	Pearson's *r*
		Depression	BASC-2	51	Pearson's *r*
Rosenberg, 2016	Parent vs. Self	Anxiety	BASC-2 PRS and SRP	20	Pearson's *r*
		Depression	BASC-2 PRS and CDI	20	Pearson's *r*

Table 30 (cont'd)

Rump, 2012	Parent vs. Self	Anxiety	SCARED	19	Pearson's *r*
		Depression	CDI	19	Pearson's *r*
Schiltz et al., 2018	Parent vs. Self	Anxiety	ASEBA CBCL and YSR	53	Pearson's *r*
Schwartz, 2010	Parent vs. Self	Anxiety	BASC-2 PRS and SRP	30	Pearson's *r*
Sterling et al., 2015	Parent vs. Self	Anxiety	RCADS and ASEBA CBCL	67	Pearson's *r*
Taylor et al., 2018	Parent vs. Self	Anxiety	BASC-2 PRS and SRP	44	Pearson's *r*
		Depression	BASC-2 PRS and SRP	44	Pearson's *r*
Ung et al., 2017	Parent vs. Teacher	Anxiety	ASEBA CBCL and TRF	32	ICC
		Depression	ASEBA CBCL and TRF	32	ICC
		Internalizing	ASEBA CBCL and TRF	32	ICC
Vickerstaff et al., 2007	Parent vs. Self	Depression	BASC PRS and SRP	22	Pearson's *r*
	Parent vs. Teacher	Depression	BASC PRS and TRF	22	Pearson's *r*
	Teacher vs. Self	Depression	BASC TRF and SRP	22	Pearson's *r*
Whitehead, 2005	Parent vs. Self	Anxiety	BASC PRS and SRP	20	Pearson's *r*

Appendix E

Table 31.

Study-by-Study Information for Hedges g Values Between Ratings of Anxiety, Depression, and Broad Internalizing

Study	Rater-pair	Construct	Measure(s)	*N* pair
Barnhill et al., 2000	Parent vs. Self	Anxiety	BASC	20
		Depression	BASC	20
	Parent vs. Teacher	Anxiety	BASC	20
		Depression	BASC	20
		Internalizing	BASC	20
	Teacher vs. Self	Anxiety	BASC	20
		Depression	BASC	20
Bellini, 2004	Parent vs. Self	Anxiety	BASC-PRS and MASC-C	41
Bitsika and Sharpley, 2015	Parent vs. Self	Anxiety	CASI	139
Bitsika et al., 2019	Parent vs. Self	Anxiety	CASI	150
		Depression	CASI	150
		Internalizing	CASI	150
Bitsika et al., 2014	Parent vs. Self	Anxiety	CASI	32
Blakeley-Smith et al., 2012	Parent vs. Self	Anxiety	SCARED	63
Boulter et al., 2014	Parent vs. Self	Anxiety	SCAS	170
Carruthers et al., 2018	Parent vs. Self	Anxiety	SCAS	38
Chalfant et al., 2007	Parent vs. Self	Anxiety	SCAS	28
	Parent vs. Teacher	Internalizing	Strengths and Difficulties Questionnaire	28

Table 31 (cont'd)

Chandler et al., 2016	Parent vs. Teacher	Anxiety	The Developmental Behavior Checklist	227
		Depression	The Developmental Behavior Checklist	277
Chiu et al., 2016	Parent vs. Self	Anxiety	ASEBA CBCL and RCADS-C	28
Chow, 2008	Parent vs. Self	Anxiety	BASC-2 and MASC-C	32
		Depression	BASC-2 and CDI-C	32
Clarke et al., 2017	Parent vs. Self	Anxiety	SCAS	14
Conaughton et al., 2017	Parent vs. Self	Anxiety	SCAS	21
Connolly, 2012	Parent vs. Teacher	Internalizing	ASEBA CBCL and TRF	71
Dauterman, 2017	Parent vs. Teacher	Internalizing	BASC-2	70
Drmic et al., 2017	Parent vs. Self	Anxiety	SCARED	35
Ellison et al., 2015	Parent vs. Teacher	Anxiety	BASC-2	67
		Depression	BASC-2	67
		Internalizing	BASC-3	67
Elzinga, 2015	Parent vs. Self	Anxiety	MASC-2	26
Foley Nicpon et al., 2010	Parent vs. Self	Anxiety	BASC-2	25
		Depression	BASC-2	25
		Internalizing	BASC-2	25
	Parent vs. Teacher	Anxiety	BASC-2	33
		Depression	BASC-2	33
		Internalizing	BASC-2	33
	Teacher vs. Self	Anxiety	BASC-2	25

Table 31 (cont'd)

		Depression	BASC-2	25
		Internalizing	BASC-2	25
Freeman, 2009	Parent vs. Self	Anxiety	BASC-2 PRS and RCMAS-C	61
		Depression	BASC-2 PRS and CDI-C	61
Hallett et al., 2013	Parent vs. Self	Anxiety	RCADS	79
Hammond and Hoffman, 2014	Parent vs. Self	Anxiety	ASI-4 and YSI-4	10
		Depression	ASI-4 and YSI-4	12
	Parent vs. Teacher	Anxiety	ASI-4	7
		Depression	ASI-4	7
	Teacher vs. Self	Anxiety	ASI-4 and YSI-4	7
		Depression	ASI-4 and YSI-4	7
Hollocks et al., 2013	Parent vs. Self	Anxiety	SCAS	38
		Depression	CDI	38
Jamison and Oeth Schuttler, 2015	Parent vs. Self	Internalizing	SSIS Internalizing	20
Jepsen et al., 2012	Parent vs. Self	Anxiety	ASEBA CBCL and YSR	44
		Depression	ASEBA CBCL and YSR	44
		Internalizing	ASEBA CBCL and YSR	44
	Parent vs Teacher	Anxiety	ASEBA CBCL and TRF	36
		Depression	ASEBA CBCL and TRF	36
		Internalizing	ASEBA CBCL and TRF	36
	Teacher vs. Self	Anxiety	ASEBA TRF and YSR	36
		Depression	ASEBA TRF and YSR	36

Table 31 (cont'd)

		Internalizing	ASEBA TRF and YSR	36
Joyce et al., 2017	Parent vs. Self	Anxiety	SCAS	13
Kaat, 2014	Parent vs. Self	Anxiety	MASC-2	43
		Depression	RCADS	41
Keith et al., 2018	Parent vs. Self	Anxiety	SCARED	26
Lane et al., 2013	Parent vs. Teacher	Anxiety	BASC-2	39
		Depression	BASC-2	39
		Internalizing	BASC-2	39
Lee, 2009	Parent vs. Self	Depression	BASC-2	30
Lopata et al., 2010	Parent vs. Self	Anxiety	BASC-2	40
		Depression	BASC-2	40
Luxford et al., 2017	Parent vs. Self	Anxiety	SCAS	18
	Parent vs. Teacher	Anxiety	SCAS-P and School Anxiety Scale	18
	Teacher vs. Self	Anxiety	The School Anxiety Scale and SCAS-C	18
Magiati et al., 2014	Parent vs. Self	Anxiety	SCAS	38
McDonald et al., 2016	Parent vs. Teacher	Anxiety	BASC-2	118
		Depression	BASC-2	118
		Internalizing	BASC-2	118
Mertens et al., 2017	Parent vs. Self	Anxiety	SCARED	22
Neil et al., 2019	Parent vs. Self	Anxiety	SCAS	19
Ooi et al., 2008	Parent vs. Self	Anxiety	SCAS	6

Table 31 (cont'd)

	Parent vs. Teacher	Anxiety	SCAS-P and The Asian Children Anxiety Scale	6
	Teacher vs. Self	Anxiety	The Asian Children Anxiety Scale and SCAS-C	6
Ooi et al., 2016	Parent vs. Self	Anxiety	SCAS	70
Peterson, 2017	Parent vs. Teacher	Internalizing	BASC-2	26
Reaven et al., 2009	Parent vs. Self	Anxiety	SCARED	10
Richdale and Baglin, 2015	Parent vs. Self	Depression	ASEBA CBCL and CDI-C-SF	17
Rodgers et al., 2016	Parent vs. Self	Anxiety	SCARED	157
Rodriguez, 2017	Parent vs. Teacher	Internalizing	ASEBA CBCL and TRF	166
Rosenberg, 2016	Parent vs. Self	Anxiety	BASC-2 PRS and MASC/MASC-2-C	20
		Depression	BASC-2 PRS and CDI-C	20
Rosen and Lerner, 2018	Parent vs. Self	Anxiety	MASC-2	51
		Depression	BASC-2	51
Rosen et al., 2019	Parent vs. Teacher	Internalizing	CASI-4R	283
Rump, 2012	Parent vs. Self	Anxiety	SCARED	19
		Depression	CDI	19
Sharpley et al., 2015	Parent vs. Self	Anxiety	CASI	16
Slavin, 2010	Parent vs. Teacher	Anxiety	BASC-2	6
		Depression	BASC-2	6
		Internalizing	BASC-2	6

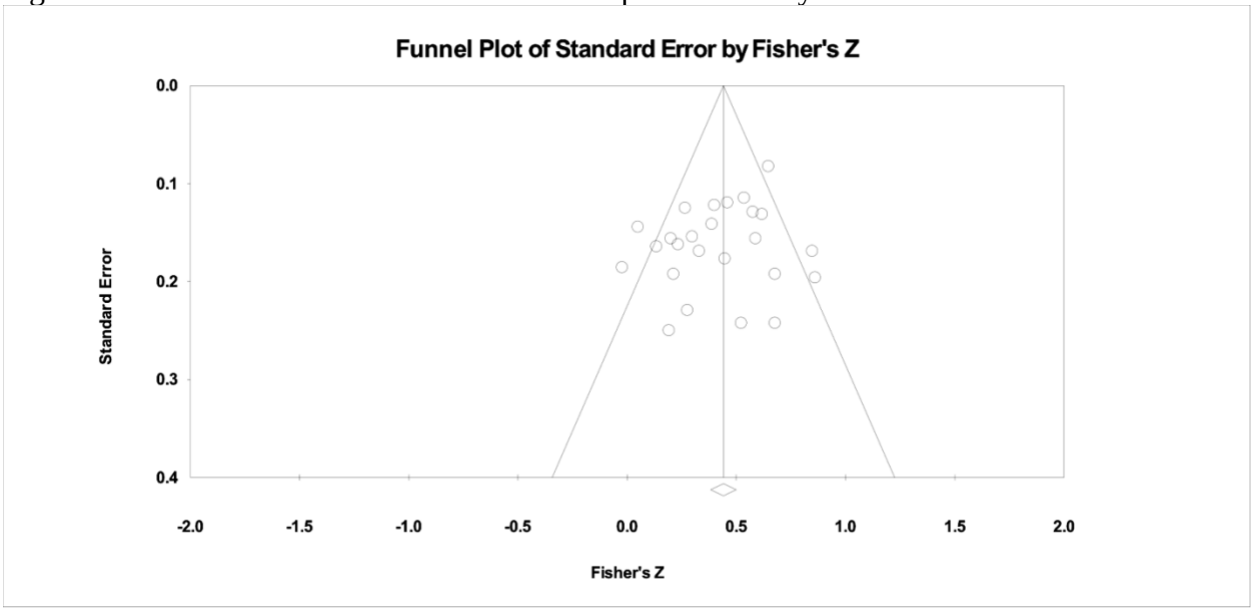
Table 31 (cont'd)

Sterling et al., 2015	Parent vs. Self	Anxiety	ASEBA CBCL and RCMAS-2-C	19
		Depression	ASEBA CBCL and RADS-2	18
Stern et al., 2014	Parent vs. Self	Anxiety	SCARED	119
Storch, 2015	Parent vs. Self	Anxiety	MASC-P and RCADS-C	16
Stratis and Lecavalier, 2017	Parent vs. Teacher	Internalizing	ASEBA CBCL and TRF	403
Taylor et al., 2018	Parent vs. Self	Anxiety	BASC-2	44
		Depression	BASC-2	44
Van Schalkwyk et al., 2018	Parent vs. Self	Anxiety	MASC-2	35
Whitehead, 2005	Parent vs. Self	Anxiety	BASC	20
		Depression	BASC	20
Wijnhoven et al., 2018	Parent vs. Self	Anxiety	SCAS	168
Wood et al., 2015	Parent vs. Self	Anxiety	MASC-P and RCADS-C	14
Wood et al., 2009	Parent vs. Self	Anxiety	MASC	14

Appendix F

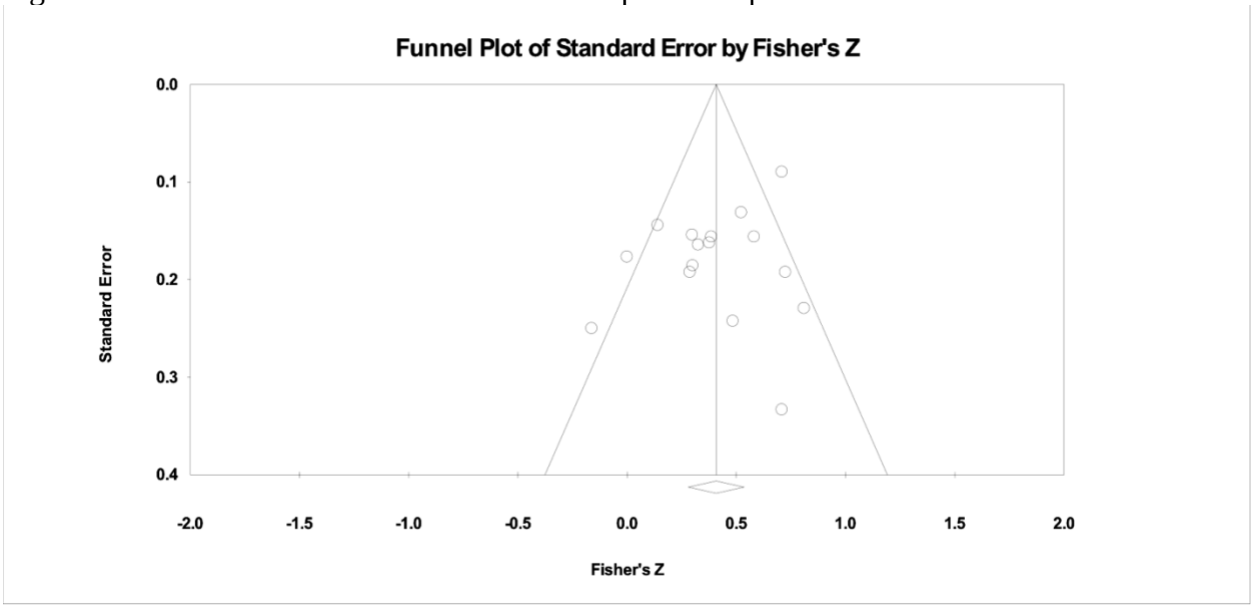
Funnel Plots with *Fail-safe N* Results for Analyses of Publication Bias

Figure 3. Correlation Between Parent vs. Self-reported Anxiety



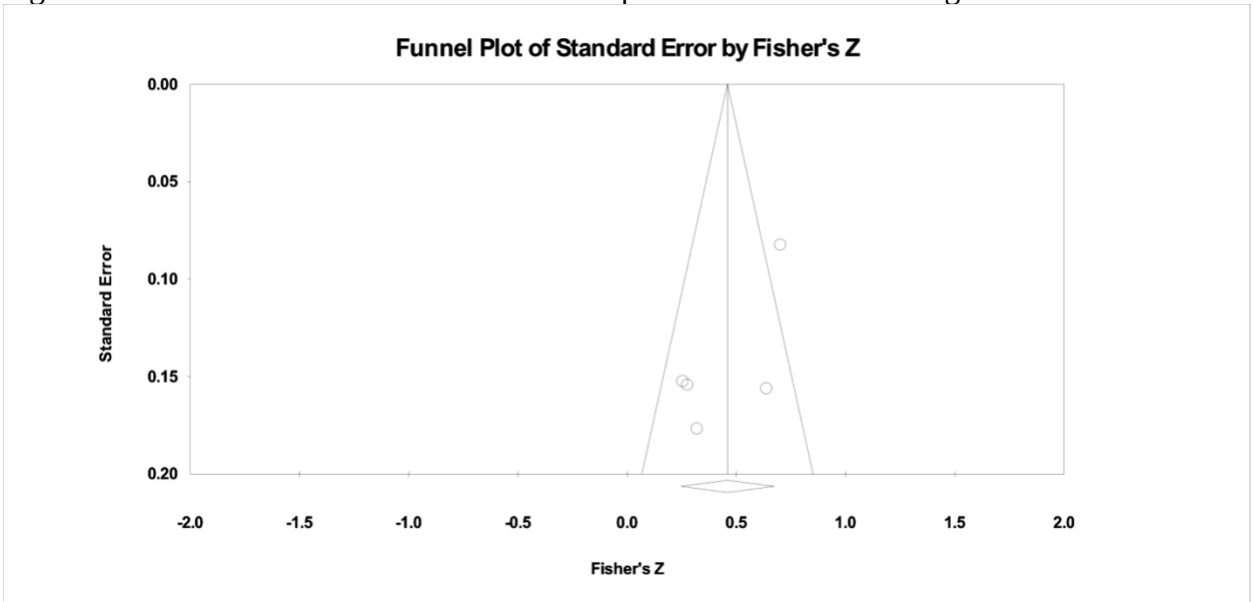
$z = 13.73$, $p < 0.001$; *Fail-safe N* = 1202; $k = 25$

Figure 4. Correlation Between Parent vs. Self-reported Depression



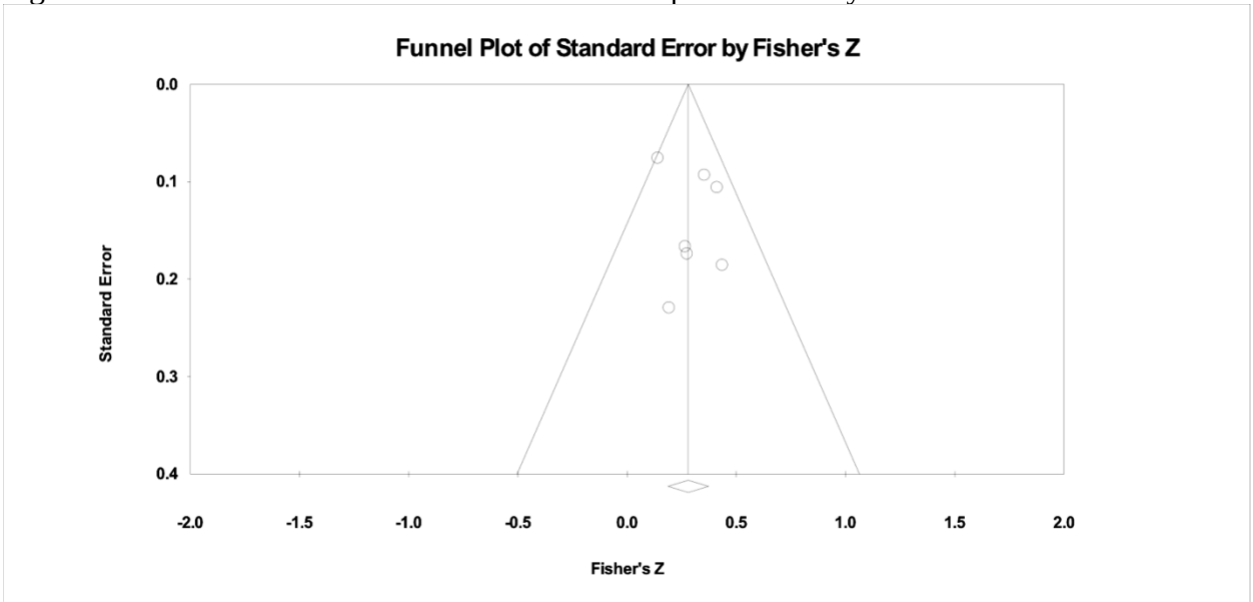
$z = 9.79$, $p < 0.001$; *Fail-safe N* = 384; $k = 16$

Figure 5. Correlation Between Parent vs. Self-reported Broad Internalizing



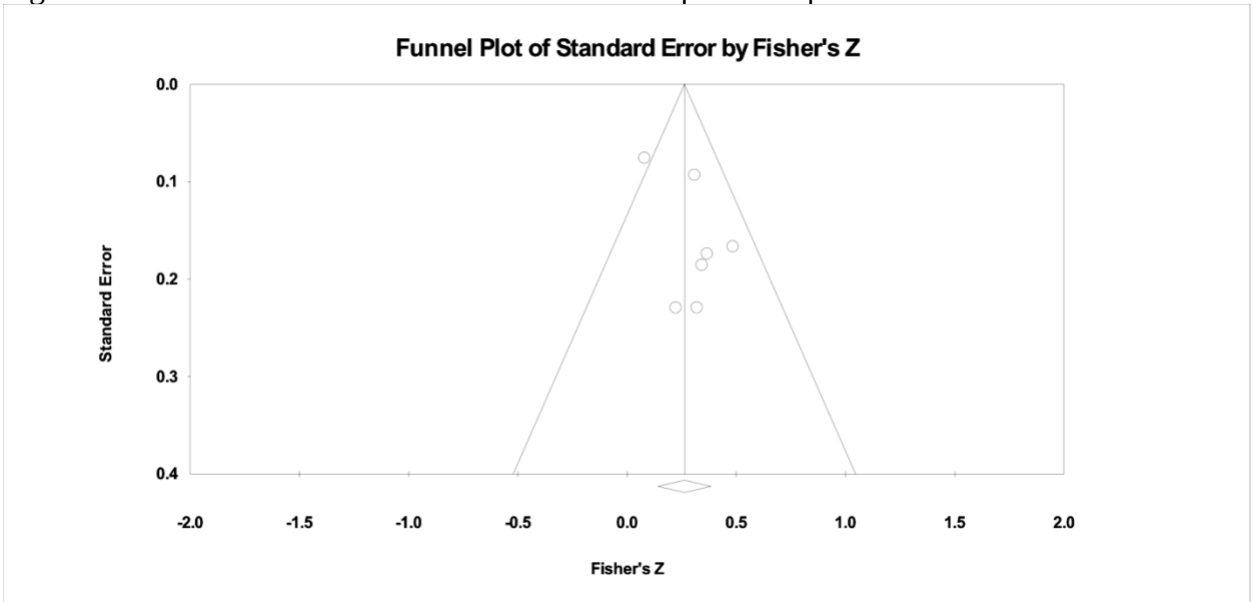
$z = 8.00$, $p < 0.001$; *Fail-safe N* = 79; $k = 5$

Figure 6. Correlation Between Parent vs. Teacher reported Anxiety



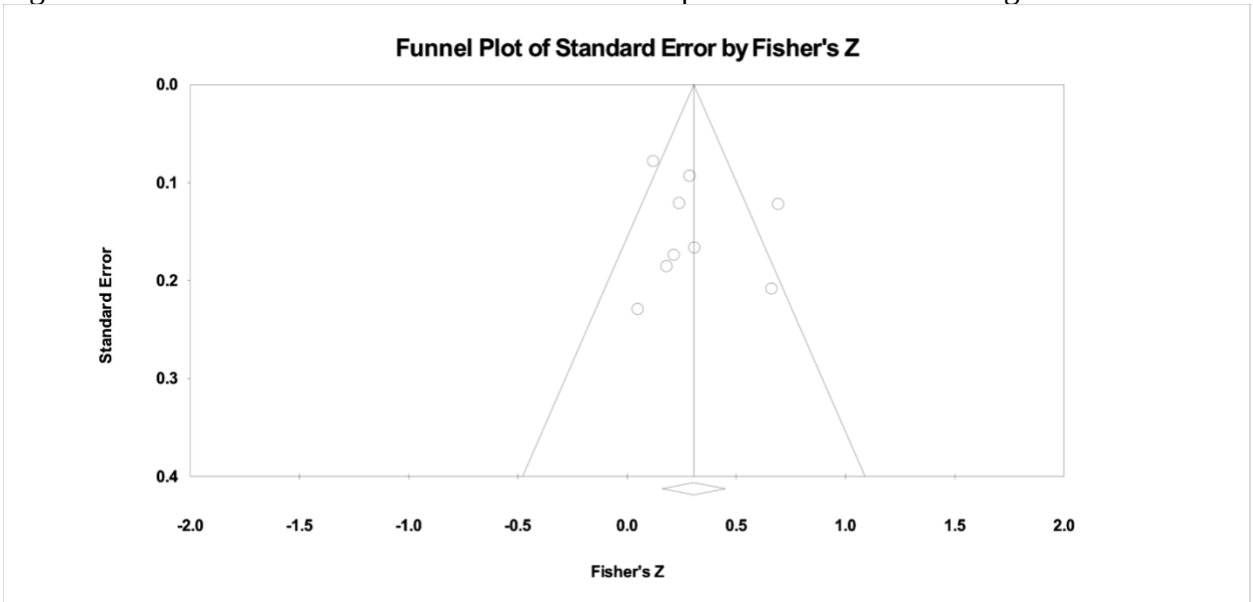
$z = 6.01$, $p < 0.001$; *Fail-safe N* = 59; $k = 7$

Figure 7. Correlation Between Parent vs. Teacher Reported Depression



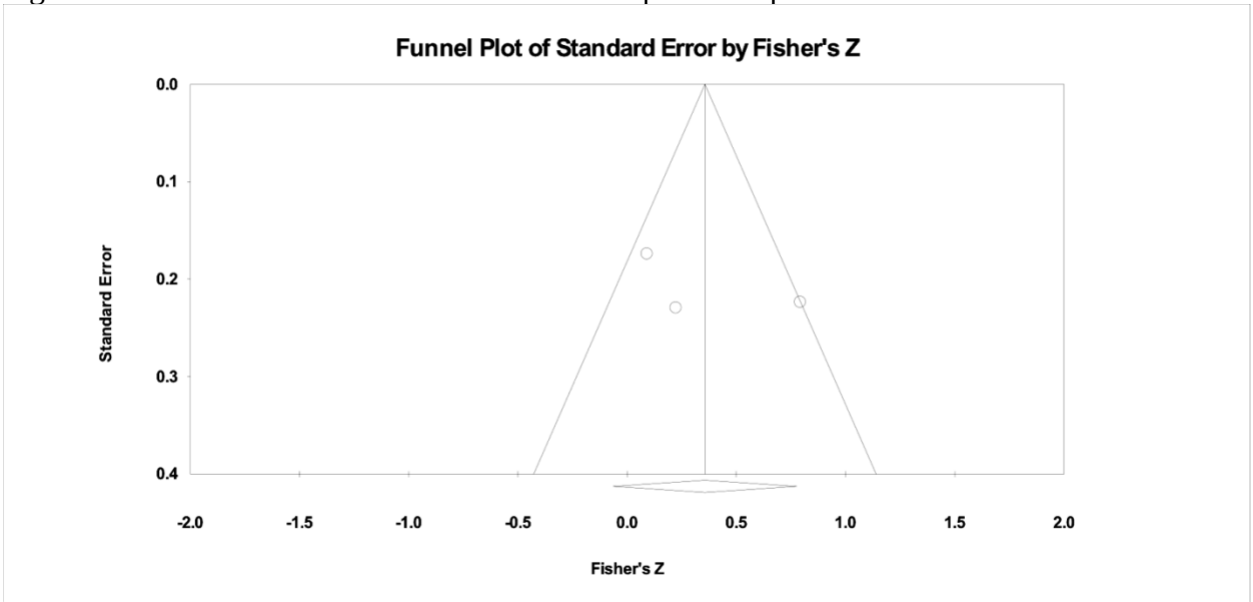
$z = 5.14$, $p < 0.001$; *Fail-safe N* = 42; $k = 7$

Figure 8. Correlation Between Parent vs. Teacher Reported Broad Internalizing



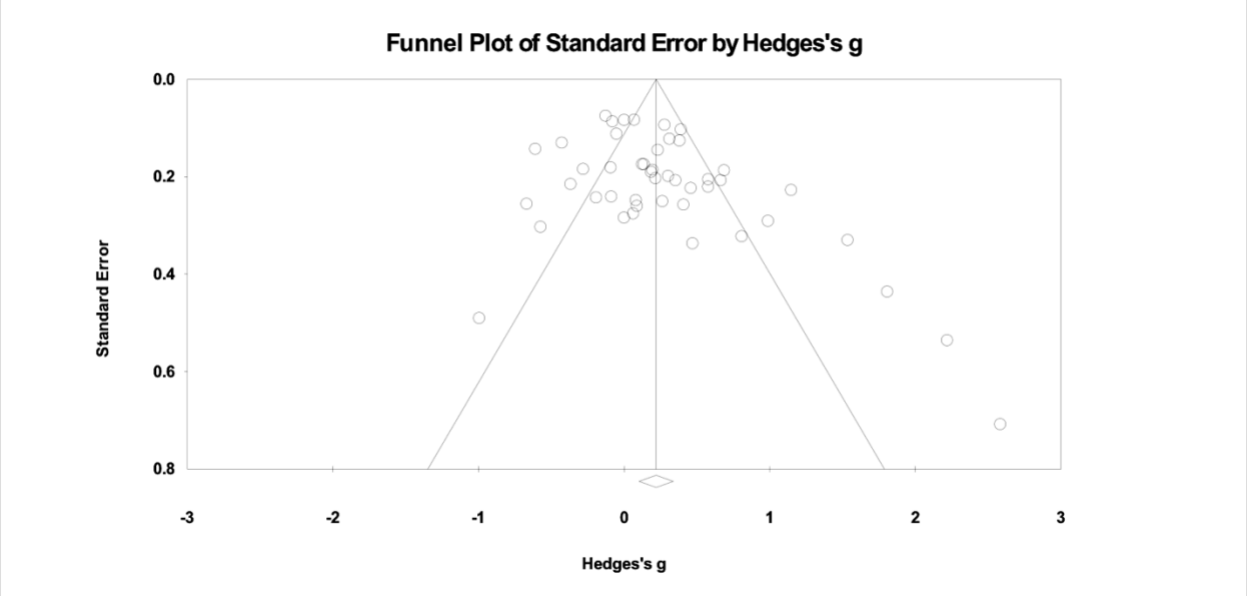
$z = 6.58$, $p < 0.001$; *Fail-safe N* = 93; $k = 9$

Figure 9. Correlation Between Teacher vs. Self-reported Depression



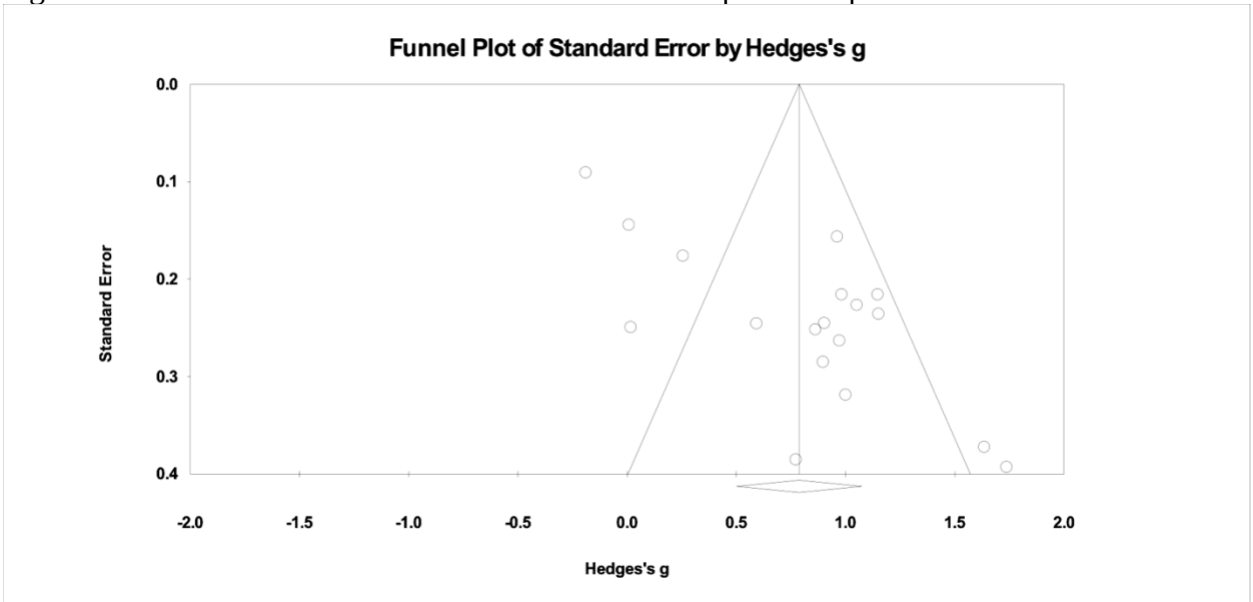
$z = 2.91$, $p = 0.003$; *Fail-safe N* = 4; $k = 3$

Figure 10. Mean Differences Between Parent vs. Self-reported Anxiety



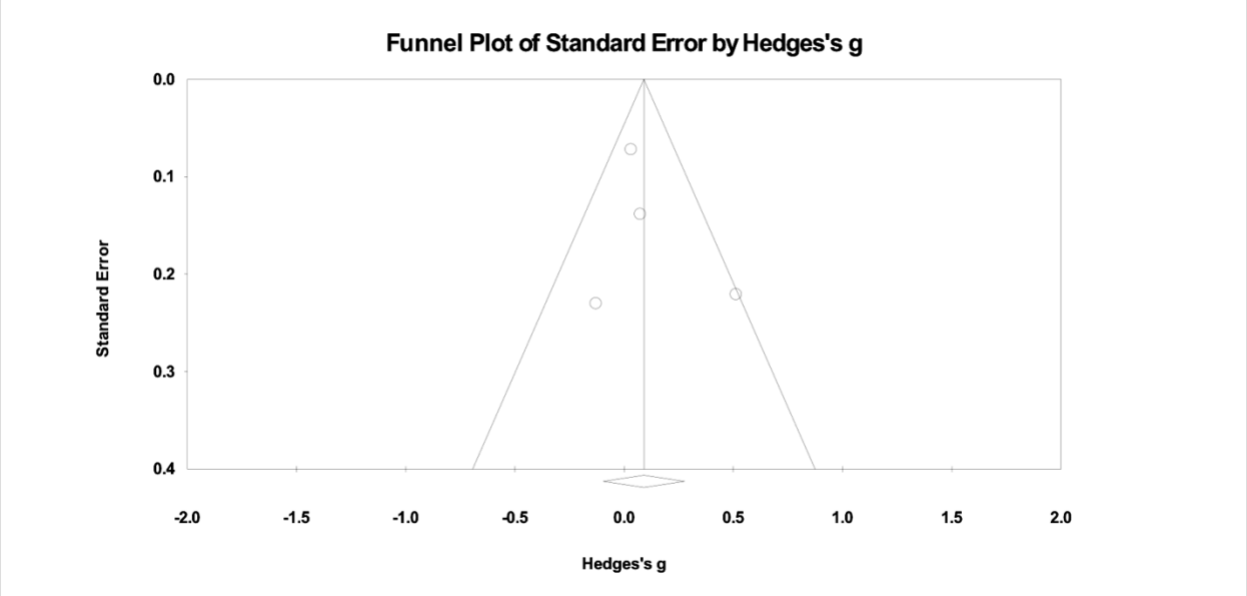
$z = 7.02$, $p < 0.001$; *Fail-safe N* = 545; $k = 46$

Figure 11. Mean Differences Between Parent vs. Self-reported Depression



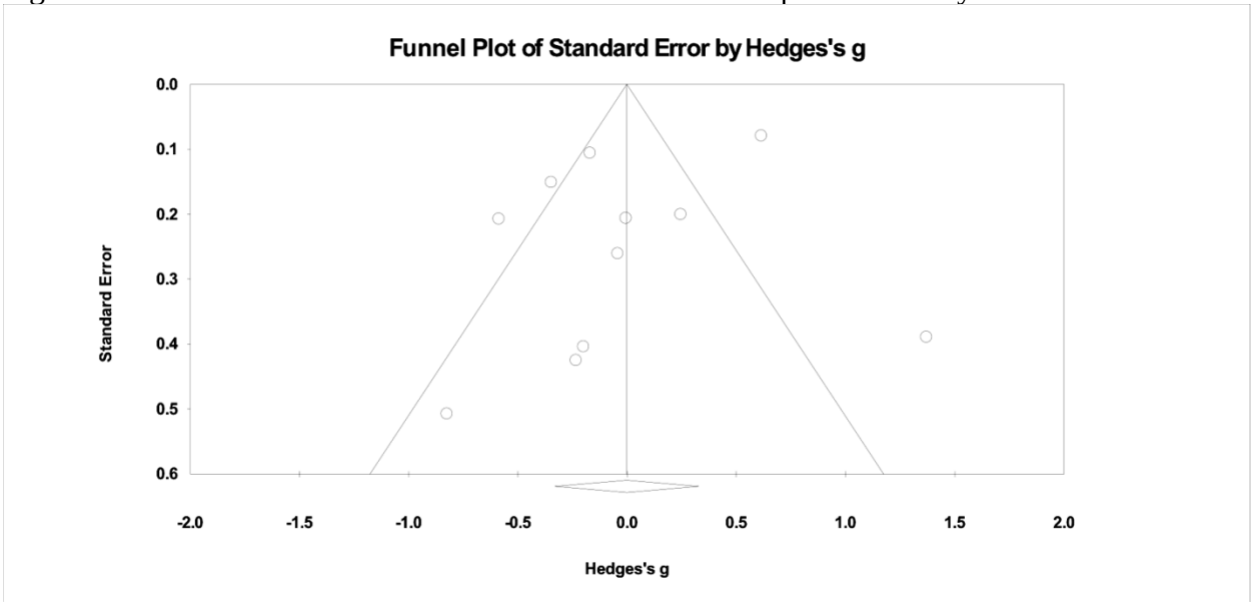
$z = 13.05$, $p < 0.001$; *Fail-safe N* = 781; $k = 18$

Figure 12. Mean Differences Between Parent vs. Self-reported Broad Internalizing



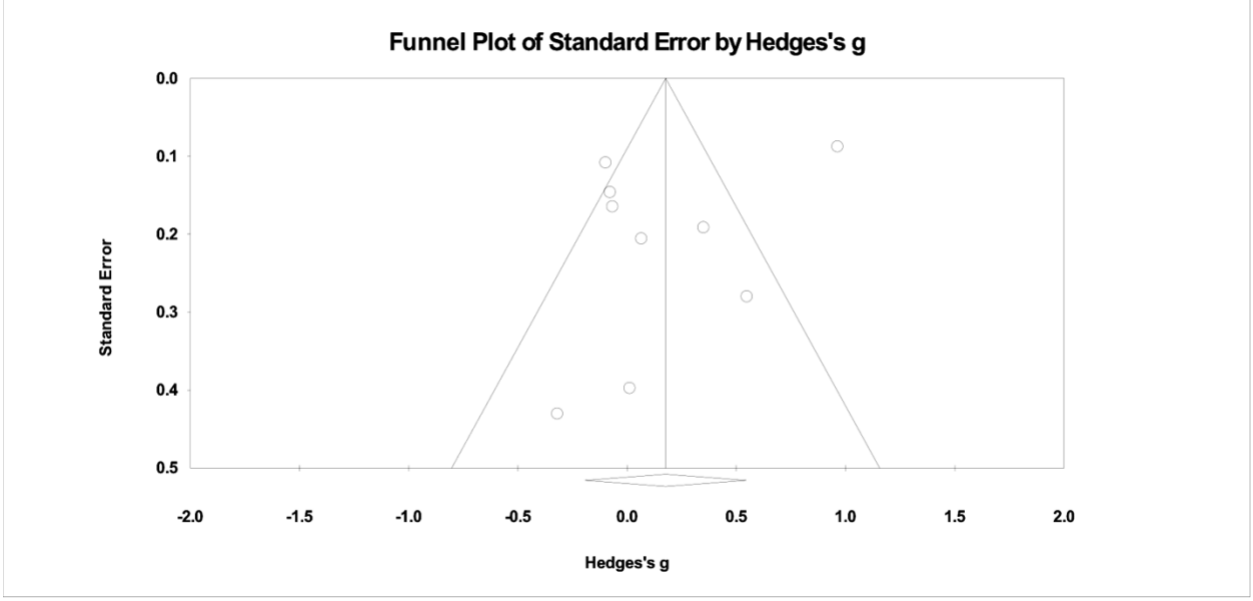
$z = 1.37$, $p = 0.17$; *Fail-safe N* = 0; $k = 4$

Figure 13. Mean Differences Between Parent vs. Teacher Reported Anxiety



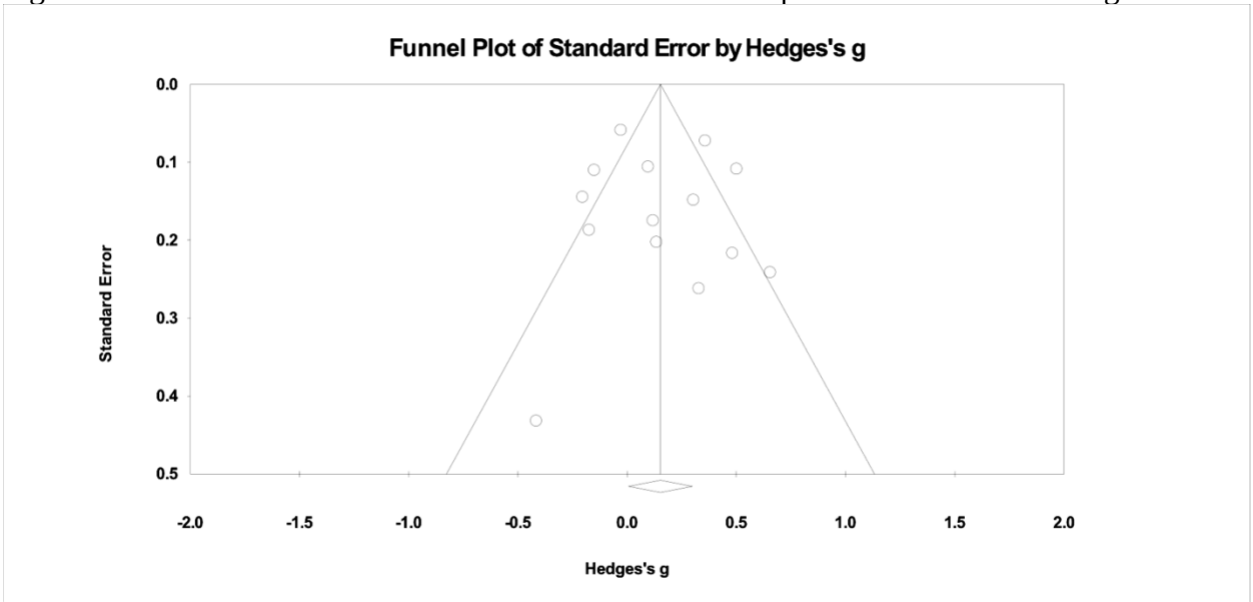
$z = 0.87$, $p = 0.379$; *Fail-safe N* = 0; $k = 11$

Figure 14. Mean Differences Between Parent vs. Teacher Reported Depression



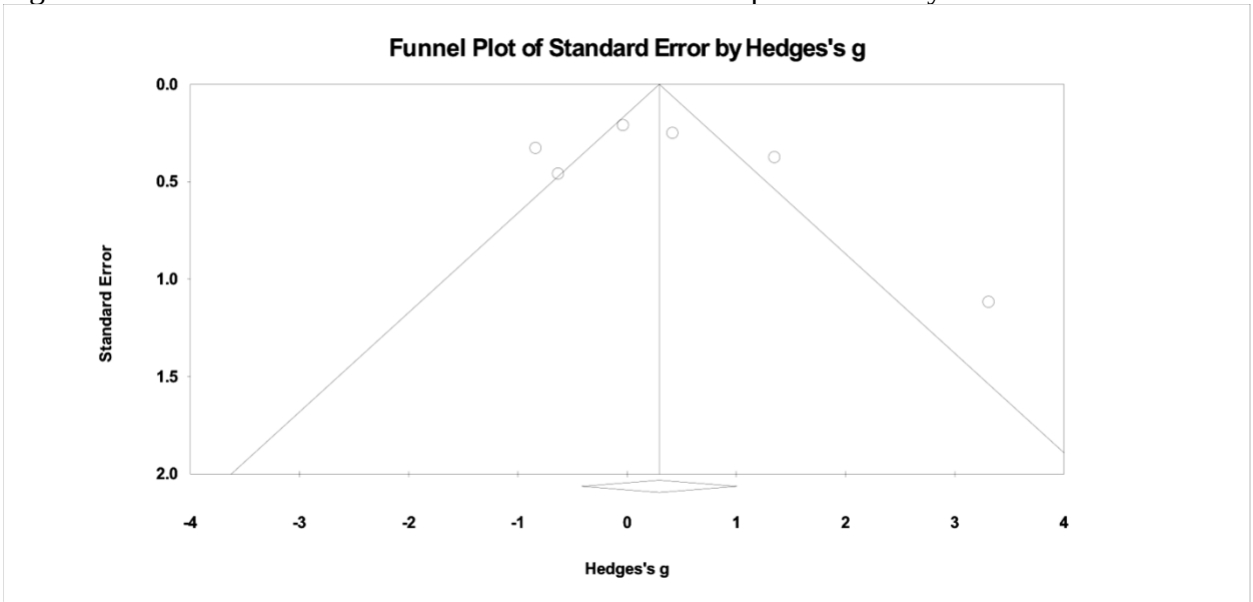
$z = 4.18$, $p < 0.001$; *Fail-safe N* = 33; $k = 9$

Figure 15. Mean Differences Between Parent vs. Teacher Reported Broad Internalizing



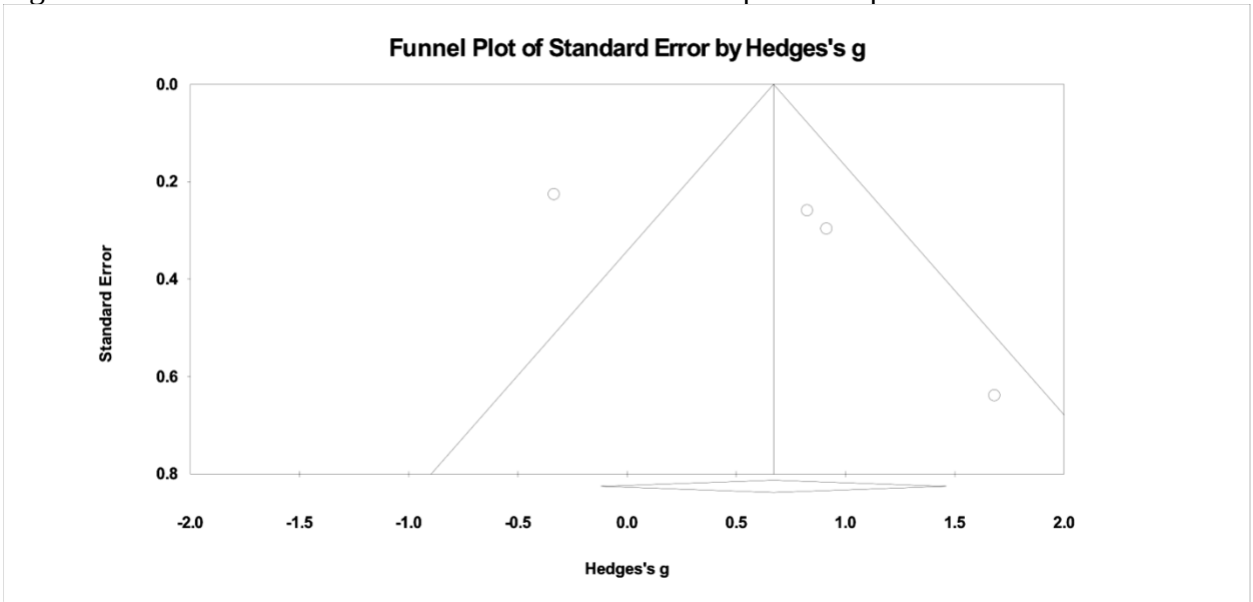
$z = 3.99$, $p < 0.001$; *Fail-safe N* = 45; $k = 9$

Figure 16. Mean Differences Between Teacher vs. Self-reported Anxiety



$z = 1.69$, $p = 0.09$; *Fail-safe N* = 0; $k = 6$

Figure 17. Mean Differences Between Teacher vs. Self-reported Depression



$z = 3.71$, $p < 0.001$; *Fail-safe N* = 11; $k = 4$

Appendix G

Scatter Plots

Figure 18. FSIQ as a Moderator Between the Correlation of Parent vs. Self-reported Anxiety

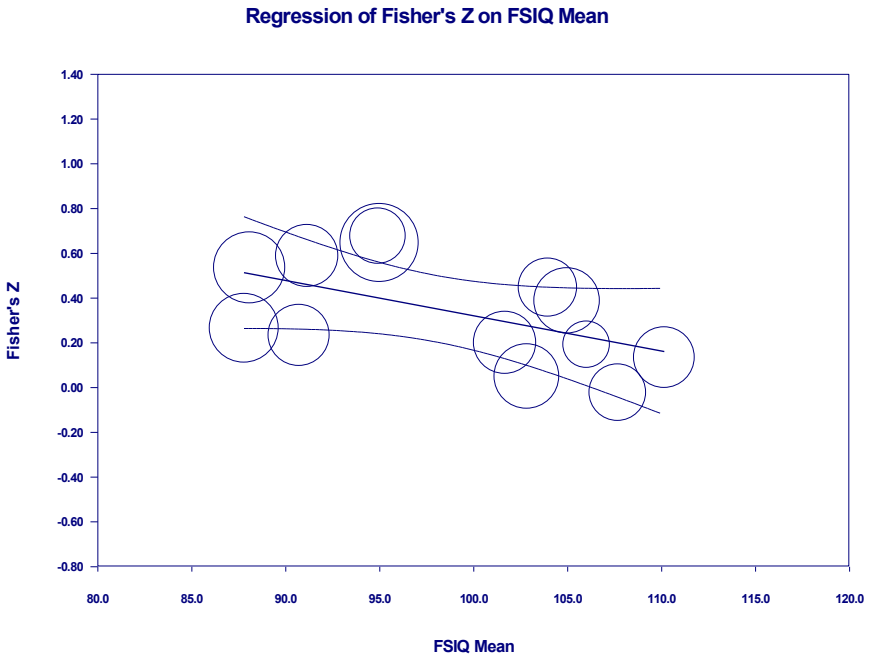


Figure 19. FSIQ as a Moderator Between the Correlation of Parent vs. Self-reported Anxiety, Follow-up Analysis: Group 1

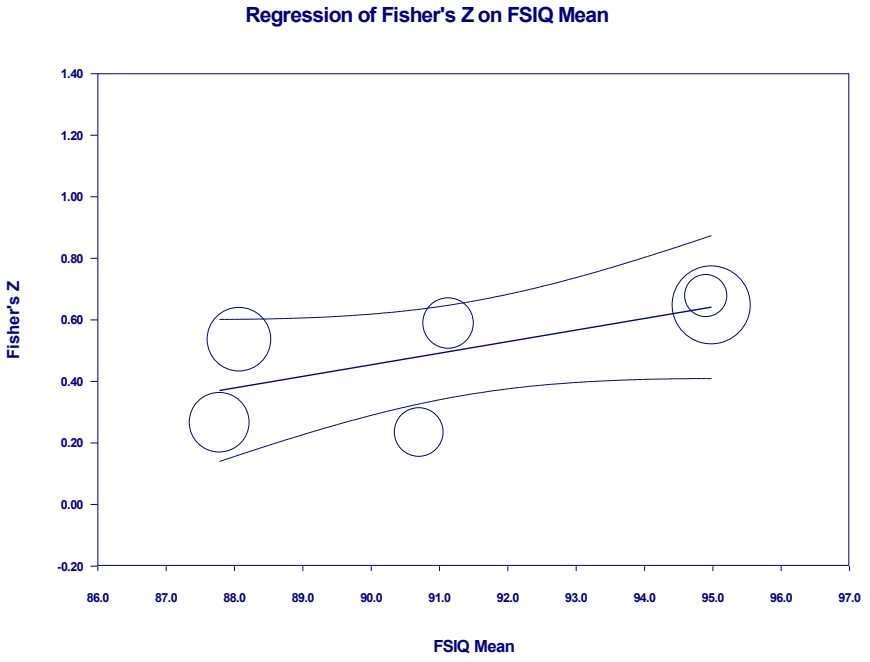


Figure 20. FSIQ as a Moderator Between the Correlation of Parent vs. Self-reported Anxiety, Follow-up Analysis: Group 2

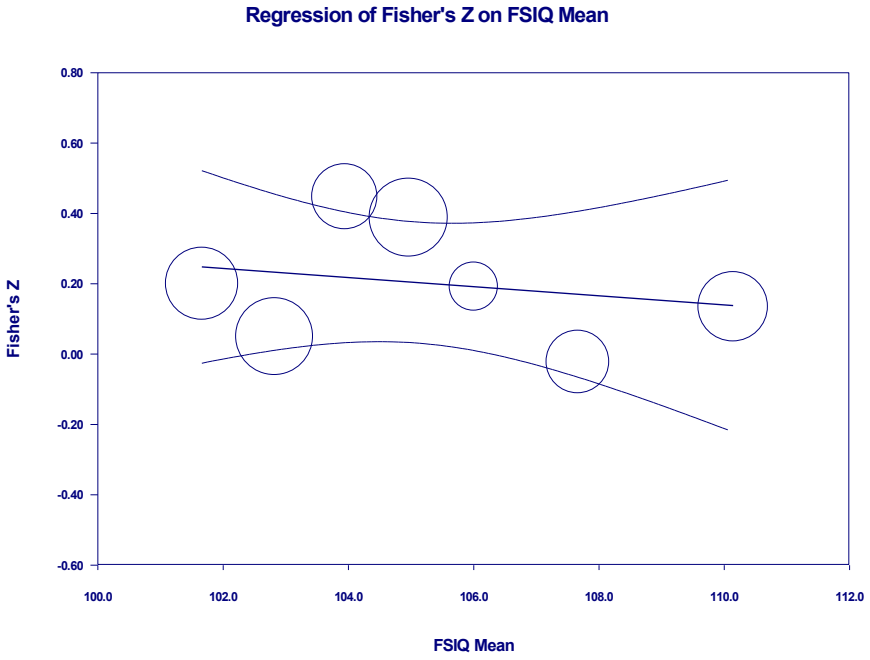


Figure 21. FSIQ as a Moderator Between the Correlation of Parent vs. Self-reported Depression

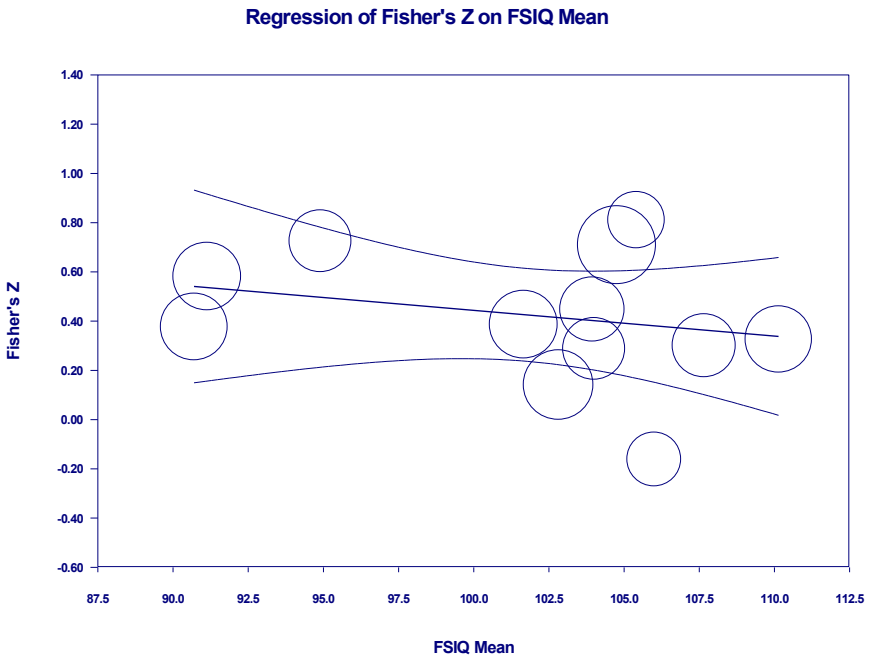


Figure 22. FSIQ as a Moderator Between the Correlation of Parent vs. Self-reported Depression, Re-run Without Outlier

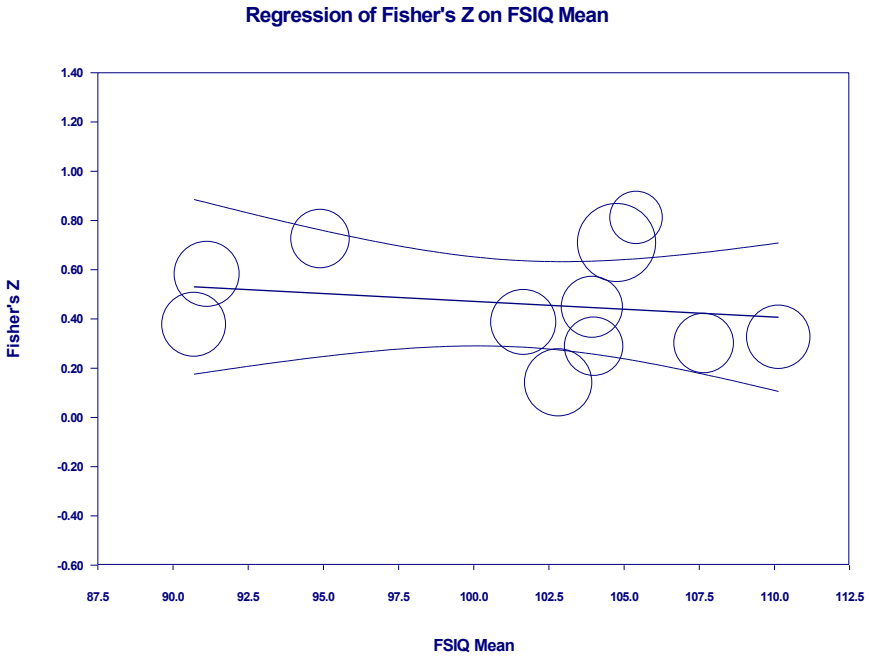


Figure 23. FSIQ as a Moderator Between the Correlation of Parent vs. Self-reported Broad Internalizing

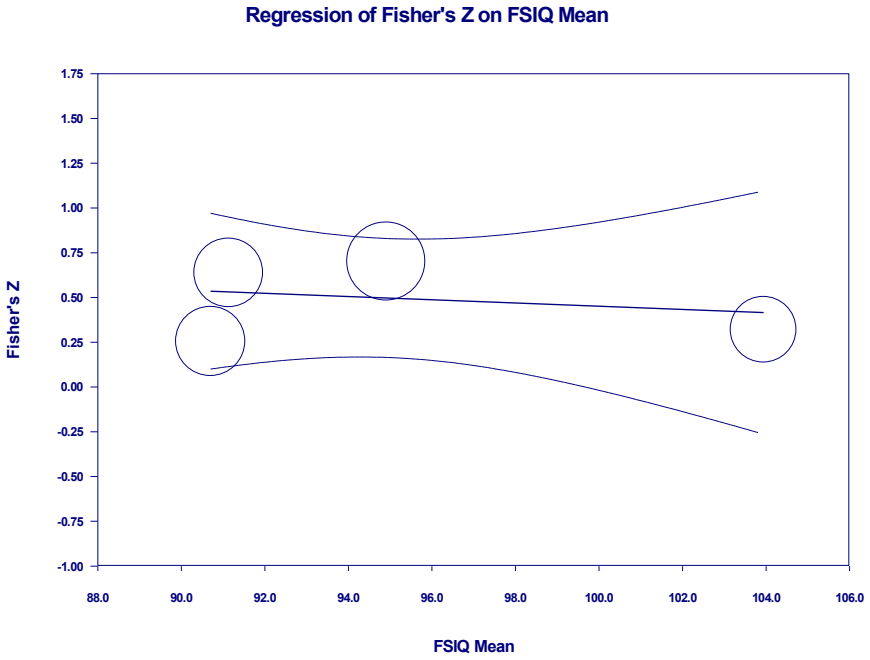


Figure 24. Age as a Moderator Between the Correlation of Parent vs. Self-reported Anxiety

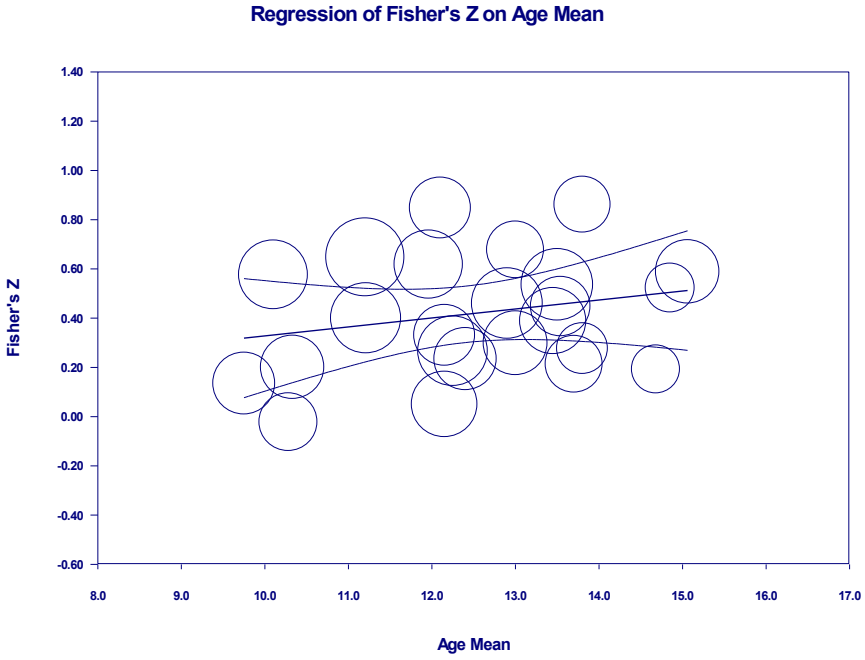


Figure 25. Age as a Moderator Between the Correlation of Parent vs. Self-reported Depression

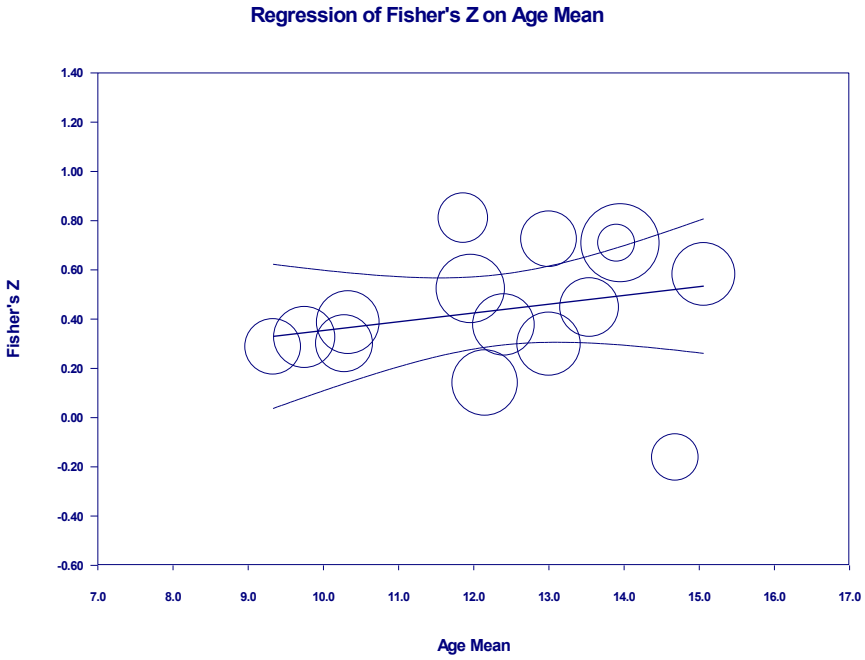


Figure 26. Age as a Moderator Between the Correlation of Parent vs. Self-reported Depression, Re-run Without Outlier

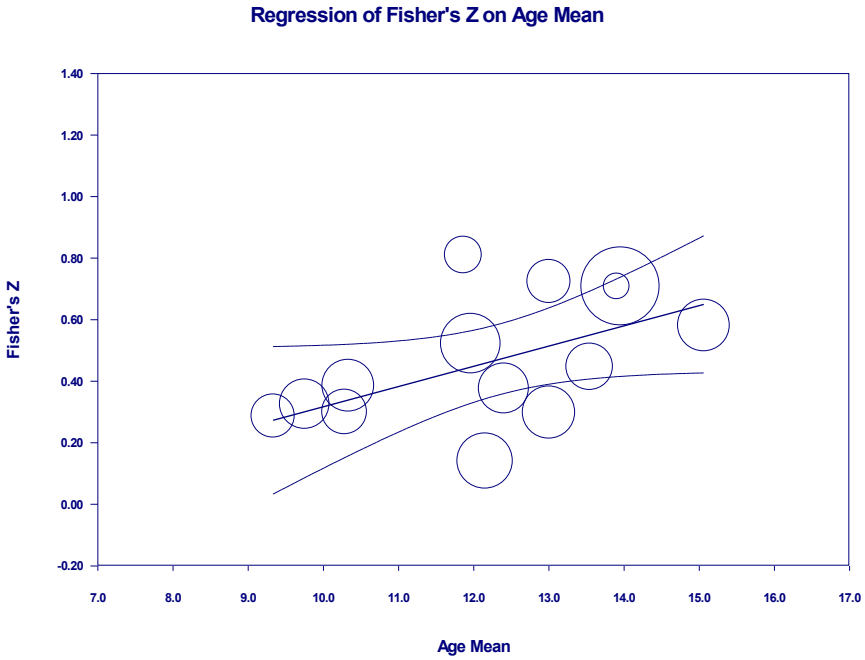


Figure 27. Age as a Moderator Between the Correlation of Parent vs. Self-reported Broad Internalizing

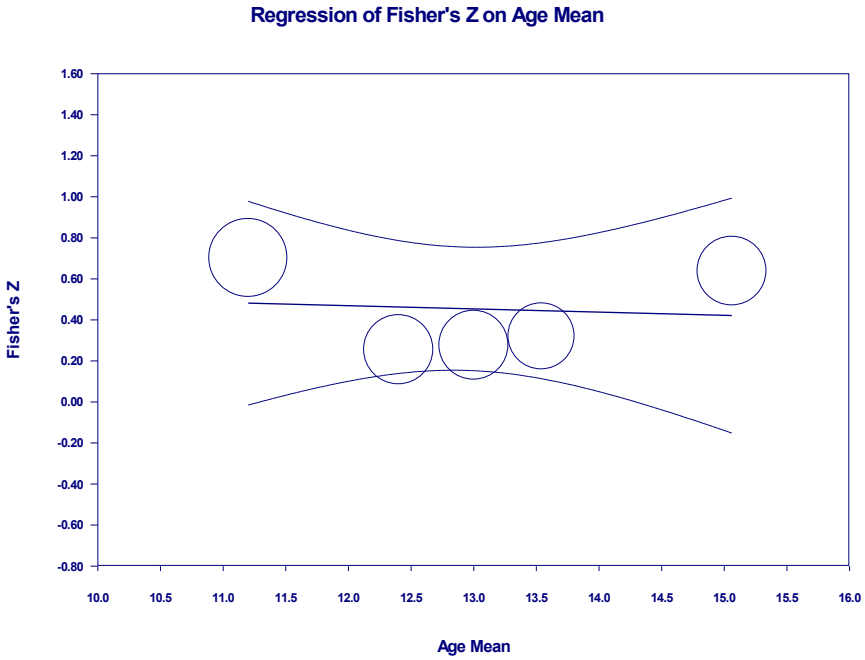


Figure 28. Method of Self-report Administration a Moderator Between the Correlation of Parent vs. Self-reported Anxiety

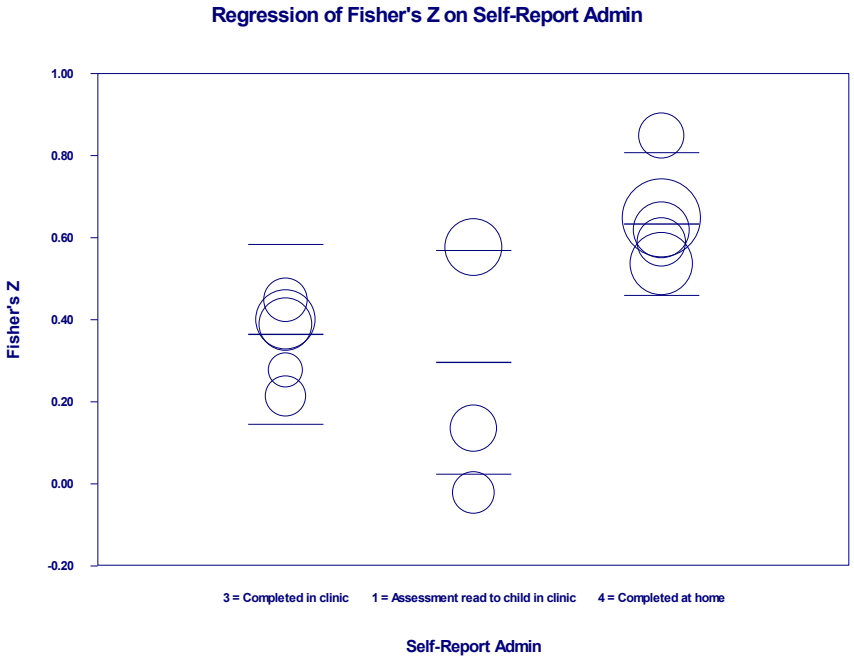


Figure 29. Method of Self-report Administration a Moderator Between the Correlation of Parent vs. Self-reported Anxiety, Re-run with Two Categories

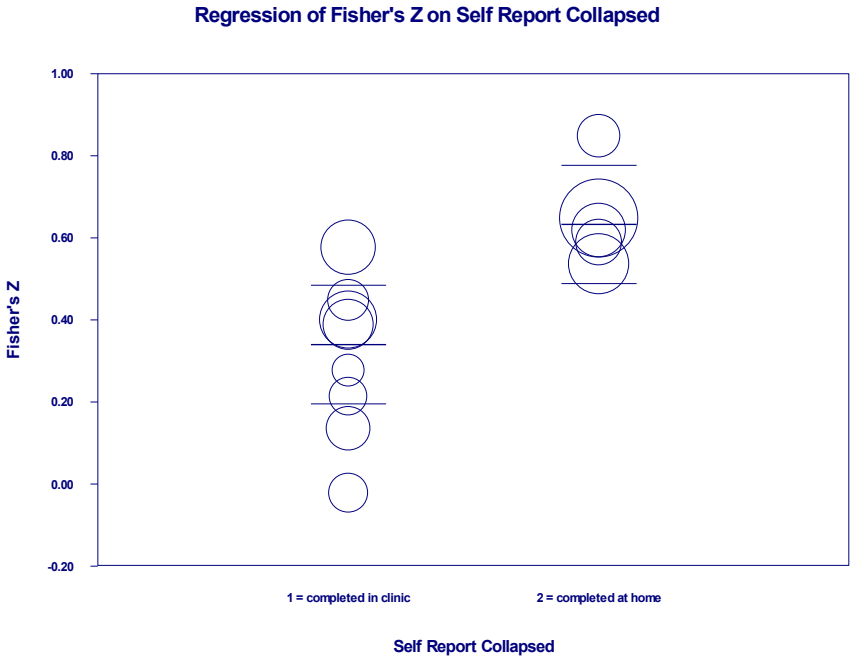


Figure 30. Method of Self-report Administration a Moderator Between the Correlation of Parent vs. Self-reported Depression

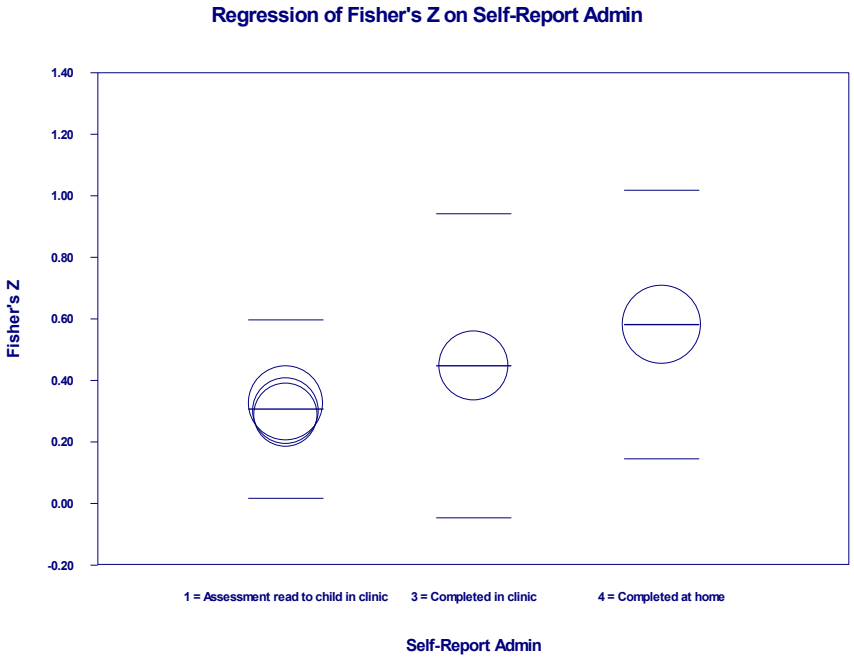


Figure 31. Method of Self-report Administration a Moderator Between the Correlation of Parent vs. Self-reported Depression, Re-run with Two Categories

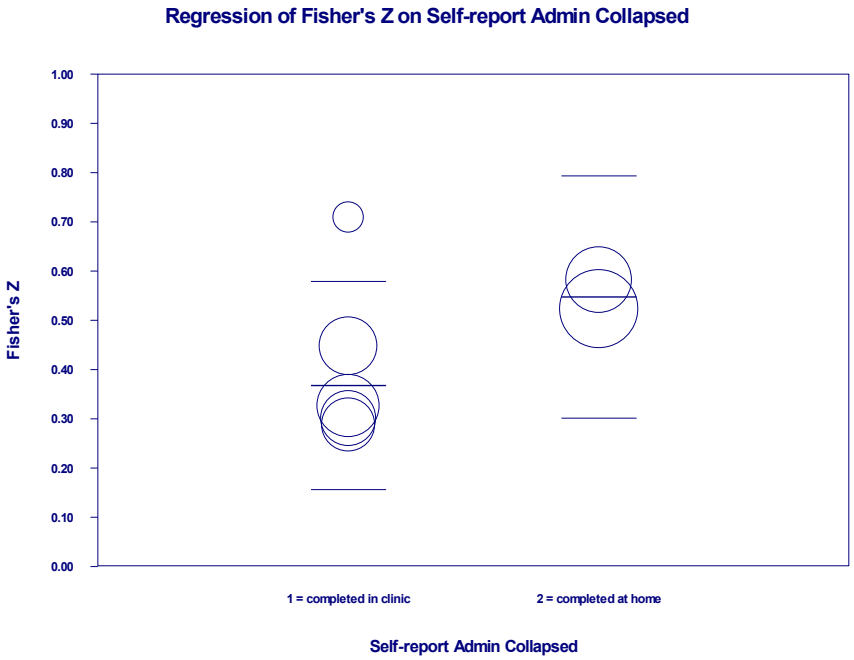


Figure 32. FSIQ as a Moderator of the Mean Differences Between Parent vs. Self-reported Anxiety

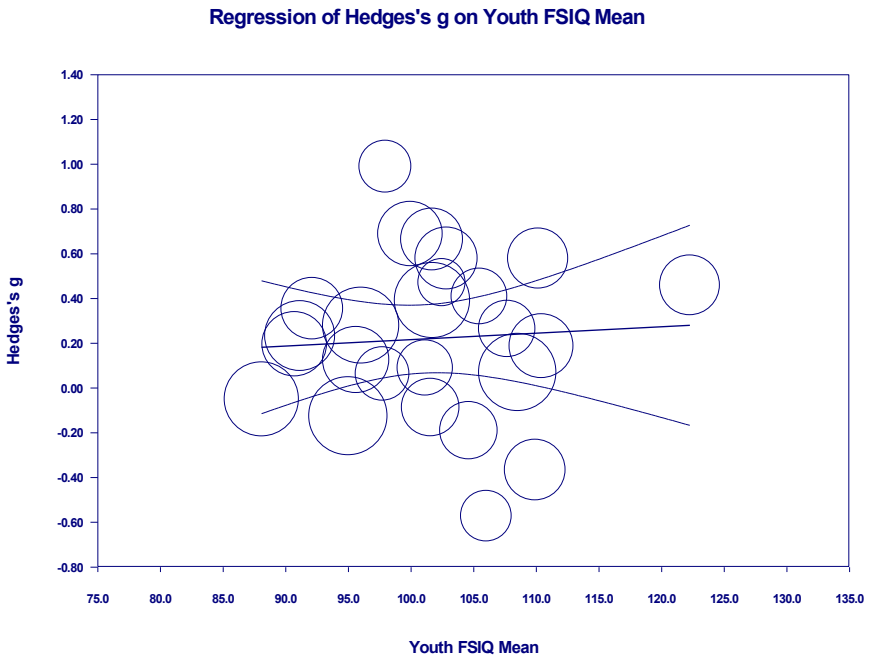


Figure 33. FSIQ as a Moderator of the Mean Differences Between Parent vs. Self-reported Depression

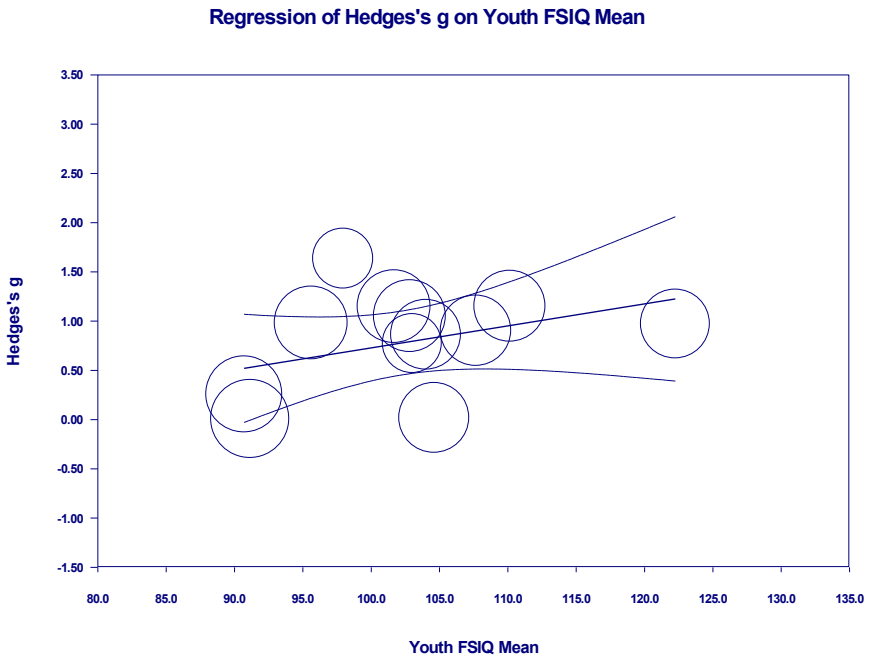


Figure 34. FSIQ as a Moderator of the Mean Differences Between Teacher vs. Self-reported Anxiety

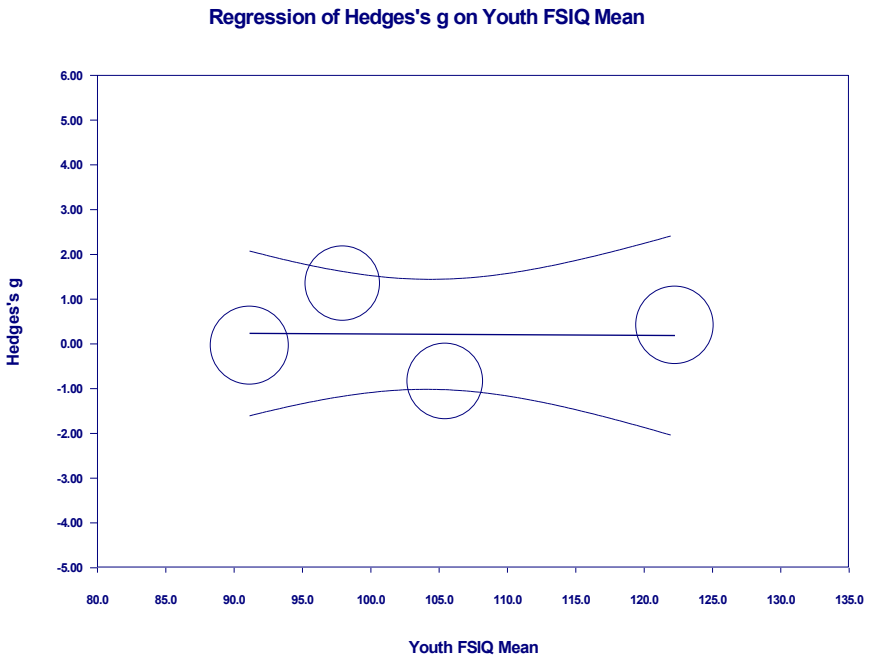


Figure 35. Age as a Moderator of the Mean Differences Between Parent vs. Self-reported Anxiety

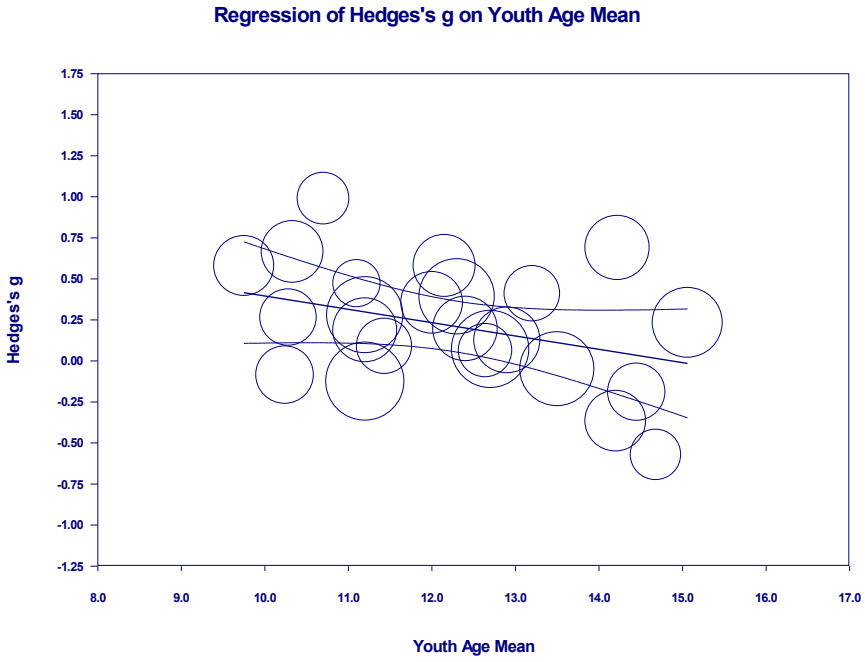


Figure 36. Age as a Moderator of the Mean Differences Between Parent vs. Self-reported Depression

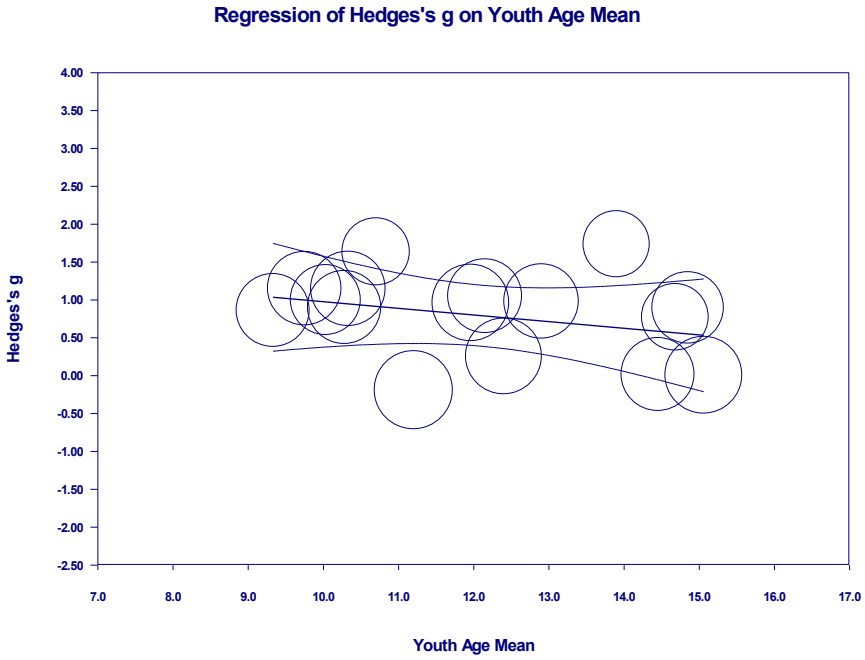


Figure 37. Age as a Moderator of the Mean Differences Between Teacher vs. Self-reported Anxiety

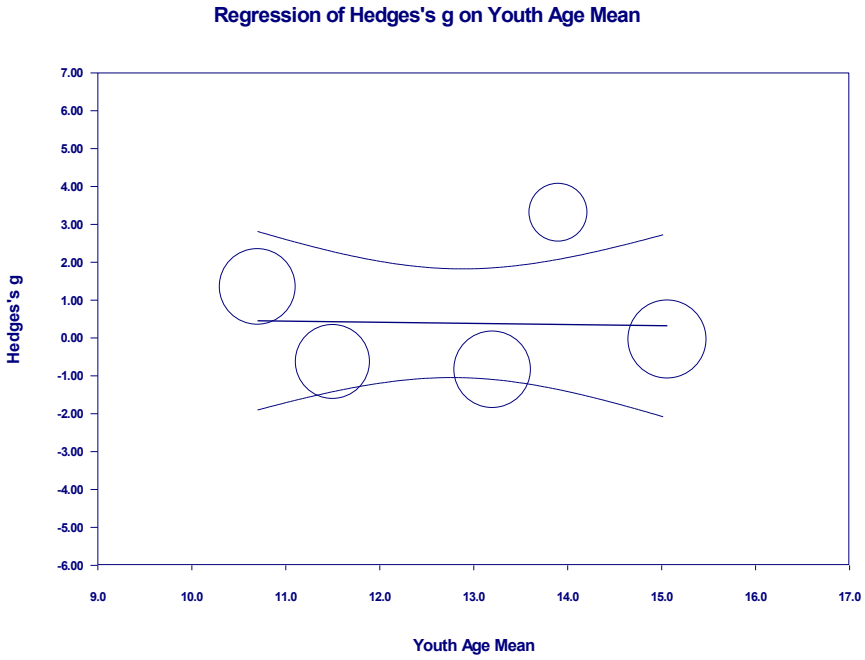


Figure 38. Method of Self-report Administration as a Moderator of the Mean Differences Between Parent vs. Self-reported Anxiety

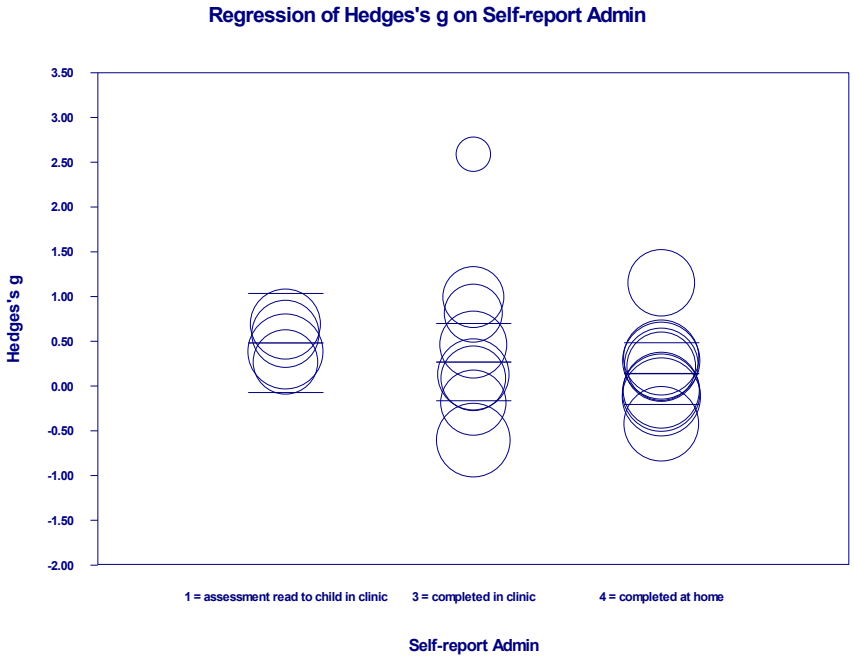


Figure 39. Method of Self-report Administration as a Moderator of the Mean Differences Between Parent vs. Self-reported Anxiety, Re-run with Two Categories

Figure 40. Method of Self-report Administration as a Moderator of the Mean Differences Between Parent vs. Self-reported Depression

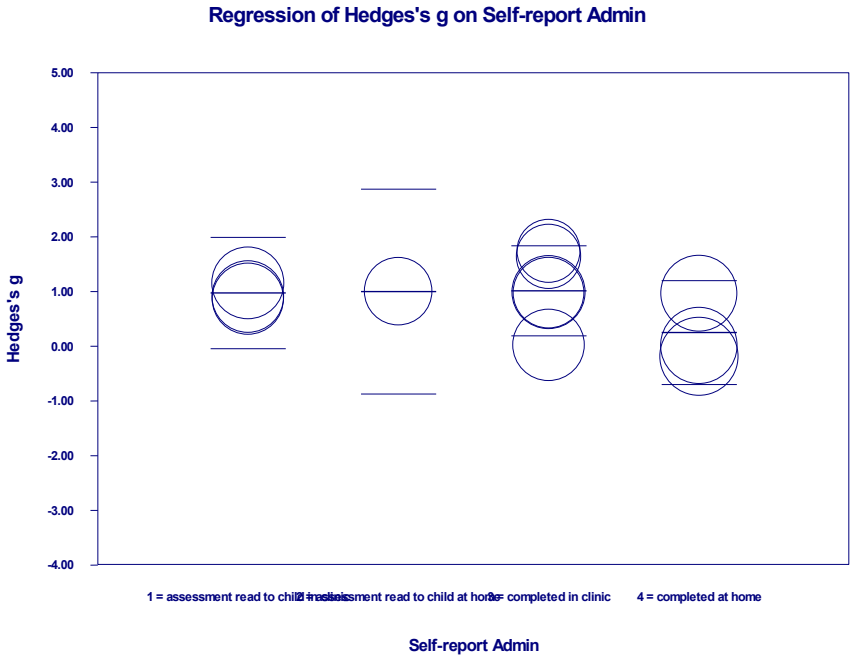
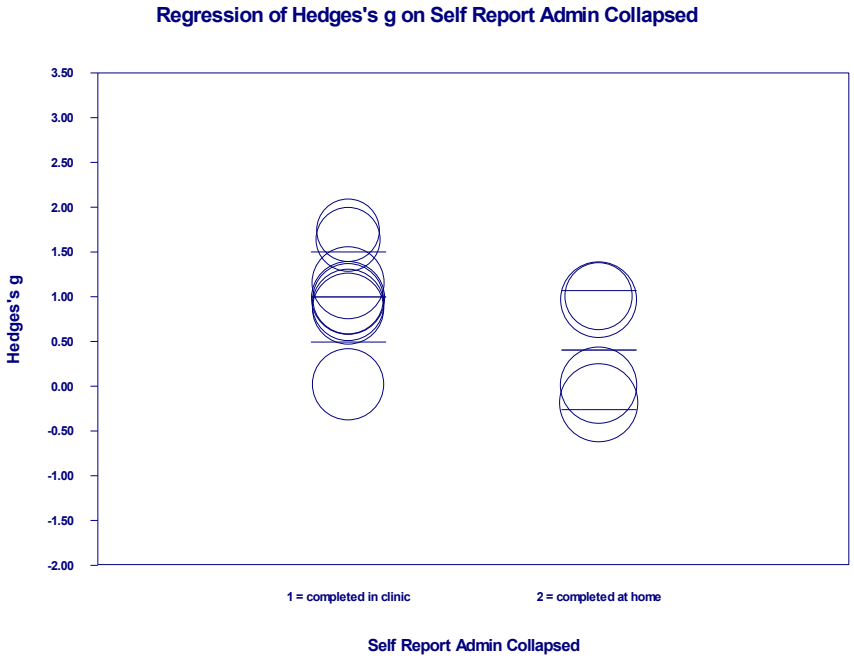


Figure 41. Method of Self-report Administration as a Moderator of the Mean Differences Between Parent vs. Self-reported Depression, Re-run with Two Categories



REFERENCES

REFERENCES

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213-232. doi: 0033-2909/87

Achenbach, T. M., & Rescorla, L. A. (2001). Manual for the ASEBA School-Age Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., Krukowski, R. A., Dumenci, L., & Ivanova, M. Y. (2005). Assessment of adult psychopathology: Meta-analyses and implications of cross-informant correlations. *Psychological Bulletin, 131*, 361-382. doi:10.1037/0033-2909.131.3.361

*Adams, D., Simpson, K., & Keen, D. (2018). School-related anxiety symptomatology in a community sample of primary-school-aged children on the autism spectrum. *Journal of School Psychology, 70*, 64–73. doi: 10.1016/j.jsp.2018.07.003

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Anderson, P. N., Skogli, E., Hovik, K., Egeland, J., & Oie, M. (2015). Associations Among Symptoms of Autism, Symptoms of Depression, and Executive Functions in Children with High-Functioning Autism: A 2 Year Follow-Up Study. *Journal of Autism and Developmental Disorders, 45*, 2497-2507. doi:10.1007/s10803-015-2415-8.

Baio, J., Wiggins, L., Christensen, D.L., Maenner, M. J., Daniels, J., Warren, Z.,…Dowling, N. F. (2018). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveillance Summaries, 67*(6), 1–23. doi: 10.15585/mmwr.ss6706a1

*Barnhill, G., Hagiwara, T. Myles, B. S., Simpson, R. L., Brick, M. L., & Griswold, D. (2000). Parent, teacher, and self-report of problem and adaptive behaviors in children and adolescents with Asperger syndrome. *Diagnostique, 25*, 147-167.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences, 6*, 248-254.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37-46. Retrieved from http://www.sciencedirect.com.proxy1.cl.msu.edu/science/article/pii/0010027785900228

*Bellini, S. (2004). Social Skill Deficits and Anxiety in High-Functioning Adolescents with Autism Spectrum Disorders. *Focus on Autism and Other Developmental Disabilities, 19*

(2), 78-86. Retrieved from http://ezproxy.msu.edu.proxy1.cl.msu.edu.proxy2.cl.msu.edu/login?url=http://search.proquest.com.proxy1.cl.msu.edu.proxy2.cl.msu.edu/docview/205057377?accountid=12598

*Bermúdez, M. O. E., Sánchez, J. J. C., del Sol, M., & Sevilla, F. (2015). Parents-Perceived and Self-Perceived Anxiety in Children with Autism Spectrum Disorder. *Educational Research and Reviews*, *10*(18), 2531–2538. Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/1773228792?accountid=12598

Bird, G. & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to emotional symptoms of autism. *Translational Psychiatry*, *3*, 1-8. doi:10.1038/tp.2013.61

Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A replication study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *38*(10), 1230–6.

Birmaher, B., Khetarpal, S., Cully, M., Brent, D., & McKenzie, S. (1995). Screen for Child Anxiety Related Emotional Disorders (SCARED).

*Bitsika, V., Arnold, W. A., & Sharpley, C. F. (2019). The role of sensory features in mediating associations between autism symptoms and anxiety in boys with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. doi: 10.1007/s10803-019-03917-1

*Bitsika, V., Sharpley, C. F., Andronicos, N. M, & Agnew, L. L. (2016). Prevalence, structure, and correlates of anxiety-depression in boys with an autism spectrum disorder. *Research in Developmental Disabilities*, *49*(50), 302-311. doi: 10.1016/j.ridd.2015.11.011

*Bitsika, V., & Sharpley, C. F. (2015). A comparison of self- vs parent reports of generalised anxiety disorder symptomatology across six age groups for boys with an ASD. *Journal of Developmental and Physical Disabilities*, *27*(2), 249–261. doi: 10.1007/s10882-014-9413-1

*Bitsika, V., Sharpley, C. F., Sweeney, J. A., & McFarlane, J. R. (2014). HPA and SAM axis responses as correlates of self- vs parental ratings of anxiety in boys with an Autistic Disorder. *Physiology & Behavior*, *127*, 1–7. doi: 10.1016/j.physbeh.2013.12.011

*Blakeley-Smith, A., Reaven, J., Ridge, K., & Hepburn, S. (2012). Parent-Child Agreement of Anxiety Symptoms in Youth with Autism Spectrum Disorders. *Research in Autism Spectrum Disorders*, *6*(2), 707–716. doi: 10.1016/j.rasd.2011.07.020

*Bohnert, A., Lieb, R., & Arola, N. (2016). More than leisure: Organized activity participation and socio-emotional adjustment among adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. doi: 10.1007/s10803-016-2783-8

Borenstein, M., Hedges, L. V., Higgins, J. P., Rothstein, H. R. (2009). Introduction to Meta-Analysis. West Sussex, United Kingdom: A John Wiley and Sons, Ltd.

*Boulter, C., Freeston, M., South, M., & Rodgers, J. (2014). Intolerance of uncertainty as a framework for understanding anxiety in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(6), 1391–1402. doi: 10.1007/s10803-013-2001-x

*Carruthers, S., Kent, R., Hollocks, M. J., & Simonoff, E. (2018). Brief report: Testing the psychometric properties of the spence children's anxiety scale (scas) and the screen for child anxiety related emotional disorders (scared) in autism spectrum disorder. *Journal of Autism and Developmental Disorders*. doi: 10.1007/s10803-018-3774-8

Centers for Disease Control and Prevention. (2014). Prevalence of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Network, 11 states, United States, 2010. *MMWR*, 63 (SS-2), 1-21. Retrieved from http://www.cdc.gov/mmwr/pdf/ss/ss6302.pdf

*Chalfant, A. M., Rapee, R., & Carroll, L. (2007). Treating Anxiety Disorders in Children with High Functioning Autism Spectrum Disorders: A Controlled Trial. *Journal of Autism and Developmental Disorders*, *37*(10), 1842–1857. doi: 10.1007/s10803-006-0318-4

*Chandler, S., Howlin, P., Simonoff, E., O'Sullivan, T., Tseng, E., Kennedy, J., … Baird, G. (2016). Emotional and behavioural problems in young children with autism spectrum disorder. *Developmental Medicine & Child Neurology*, *58*(2), 202–208. doi: 10.1111/dmcn.12830

*Chiu, T. A., Anagnostou, E., Brian, J., Chau, T., & Kushki, A. (2016). Specificity of autonomic arousal to anxiety in children with autism spectrum disorder. *Autism Research*, *9*(4), 491–501. doi: 10.1002/aur.1528

Chorpita, B. F., Moffitt, C., & Gray, J. (2005). Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behaviour Research and Therapy, 43*, 309-322.

Chorpita, B. F., Yim, L. M., Moffitt, C. E., Umemoto L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A Revised Child Anxiety and Depression Scale. *Behaviour Research and Therapy, 38*, 835-855.

*Chow, S. Y. (2008). Depression and anxiety in children with High -Functioning Autism Spectrum Disorders: Examination of clinical symptoms, assessment methods and source differences (State University of New York at Buffalo). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/89303240?accountid=12598

Christensen, D. L., Baio, J., Braun, K. V., Bilder, D. B., Charles, J., Constantino, J. N., Daniels, J.,…&Yeargin-Allsopp, M. (2016). Prevalence and characteristics of Autism Spectrum

Disorder among children aged 8-years. *Autism and Developmental Disabilities Monitoring Network.* doi: 10.15585/mmwr.ss6503a1

*Clarke, C., Hill, V., & Charman, T. (2017). School Based Cognitive Behavioural Therapy Targeting Anxiety in Children with Autistic Spectrum Disorder: A Quasi-Experimental Randomised Controlled Trail Incorporating a Mixed Methods Approach. *Journal of Autism and Developmental Disorders*, *47*(12), 3883–3895. doi: 10.1007/s10803-016-2801-x

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. Routledge.

*Conaughton, R. J., Donovan, C. L., & March, S. (2017). Efficacy of an Internet-based CBT program for children with comorbid high functioning autism spectrum disorder and anxiety: A randomised controlled trial. *Journal of Affective Disorders*, *218*, 260–268. doi: 10.1016/j.jad.2017.04.032

*Connolly, A. A. (2012). The relationship of primary caregiver perceptions of language and behavioral levels of children with autism to primary caregiver stress and ratings of family climate (City University of New York). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/906492811?accountid=12598

Cooper, H. (2017). Research synthesis and meta-analysis: A step-by-step approach (5th ed.). Los Angeles, CA: SAGE Publications Inc.

Corey, D. M., Dulap, W. P., & Burke, M. J. (1998). Averaging correlations: Expected values and bias in combined Pearson *r*s and Fisher's *z* transformations. *The Journal of General Psychology, 125*, 245-261.

*Dauterman, H. A. (2017). Adaptive functioning deficits and internalizing problems in young children with autism spectrum disorders (Seattle Pacific University). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/1895310758?accountid=12598

Davidsson, M., Hult, N., Gillberg, C., Särneö, C., Gillberg, C., & Billstedt, E. (2017). Anxiety and depression in adolescents with ADHD and autism spectrum disorders; correlation between parent- and self-reports and with attention and adaptive functioning. *Nordic Journal of Psychiatry, 71*(8), 614-620. doi: 10.1080/08039488.2017.1367840

De Bruin, E. I., Ferdinand, R. F., Meester, S., de Nijs, P. F. A., & Verheij, F. (2007). Journal of Autism and Developmental Disorders, 37, 877-886.

De Los Reyes, A., & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment*, 16, 330-334. doi:10.1037/1040-3590.16.3.330

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and

recommendations for further study. *Psychological Bulletin, 113*(4), 483-509. doi: 10.1037/0033-2909.131.4.483

*Drmic, I. E., Aljunied, M., & Reaven, J. (2017). Feasibility, acceptability and preliminary treatment outcomes in a school-based CBT intervention program for adolescents with ASD and anxiety in Singapore. *Journal of Autism and Developmental Disorders, 47*(12), 3909–3929. doi: 10.1007/s10803-016-3007-y

Duhig, A. M., Renk, K., Epstein, M. K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. Clinical Psychology: Science and Practice, 7, 435–453.

Durbin, E. C. (2010). Validity of young children's self-reports of their emotion in response to structured laboratory tasks. *Emotion, 10*(4), 519-535. doi: 10.1037/a0019008

Duval, S & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463. doi: 10.111/j.006-341x2000.00455.x

Ebesutani, C., Bernstein, A., Martinez, J. I., Chorpita, B. F., & Weisz, J. R. (2011). The Youth Self Report: Applicability and validity across younger and older youths. *Journal of Clinical Child & Adolescent Psychology, 40*(2), 338-346. doi: 10.1080/15374416.2011.546041

*Ellison, K. (2015). Comparisons: BASC-2 Parent and Teacher Reports for Children on the DSM-5 Autism Spectrum (Eastern Kentucky University). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/1721391902?accoun acc=12598

*Elzinga, N. M. (2015). Victimization among youth with autism spectrum disorder: An examination of youth anxiety and the role of social support (Northern Illinois University). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/1765406975?accoun tid=12598

Emerson, E., Felce, D., & Stancliffe, R. J. (2013). Issues concerning self-report data and population-based data sets involving people with intellectual disabiliites. *Intellectual and Developmental Disabilities, 51* (5), 333-348. doi: https://doi.org/10.1352/1934-9556-51.5.333

*Farrugia, S., & Hudson, J. (2006). Anxiety in adolescents with Asperger syndrome: Negative thoughts, behavioral problems, and life interference. *Focus on Autism and Other Developmental Disabilities, 21,* 25-35.

*Foley Nicpon, M., Doobay, A. F., Assouline, S. G., Nicpon, M. F., Doobay, A. F., & Assouline, S. G. (2010). Parent, teacher, and self perceptions of psychosocial functioning in

intellectually gifted children and adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *40*(8), 1028–1038. doi: 10.1007/s10803-010-0952-8

*Freeman, M. (2009). Examination of the Asperger syndrome profile in children and adolescents: Behaviour, mental health and temperament (University of Guelph (Canada)). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/304890300?accountia=12598

Frick, P. J., Silverthorn P., & Evans, C. (1994). Assessment of childhood anxiety using structured interviews: patterns of agreement among informants and association with maternal anxiety. *Psychological Assessment, 6*, 372–379.

Gadow, K. D., DeVincent, C. J., Pomeroy, J., & Azazian, A. (2004). Psychiatric symptoms in preschool children with PDD and clinic comparison samples. *Journal of Autism and Developmental Disorders, 34*, 379-393.

Ghaziuddin, M., Ghaziuddin, N., & Greden, J. (2002). Depression in persons with autism: Implications for research and clinical care. *Journal of Autism and Developmental Disorders, 32*(4), 299-306. doi: 0162-3257/02/0800-0299

Goodwin, A., Matthews, N. L., & Smith, C. J. (2017). The effects of early language on age at diagnosis and functioning at school age in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*, 2176-2188. doi: 10.1007/s10803-017-3133-1

Gray, L. B., Dubin-Rhodin, A., Weller, R. A., & Weller, E. B. (2009). Assessment of depression in children and adolescents. *Current Psychiatry Reports, 11*, 106-113.

Hagopian, L., & Jennet, H. (2014). Behavioral assessment and treatment of anxiety for those with autism spectrum disorder. In J. Matson (Ed.), *Handbook of Autism and Anxiety* (pp. 155-169). Switzerland: Springer.

*Hallett, Victoria, Ronald, A., & Colvert, Emma, Ames, Catherine Woodhouse, Emma Lietz, Stephanie Garnett, Tracy Gillan, Nicola Rijsdijk, Fruhling Scahill, Lawrence Bolton, Patrick Happé, F. (2013). Exploring anxiety symptoms in a large-scale twin study of children with autism spectrum disorders, their co-twins and controls. *Journal of Child Psychology and Psychiatry*, *54*(11), 1176–1185. doi: 10.1111/jcpp.12068

*Hammond, R. K., & Hoffman, J. M. (2014). Adolescents with high-functioning autism: An investigation of comorbid anxiety and depression. *Journal of Mental Health Research in Intellectual Disabilities*, *7*(3), 246–263. doi: 10.1080/19315864.2013.843223

Happe, F. (2003). Theory of mind and the self. *Annals of the New York Academy of Sciences, 1001*, 134-144.

Heaton, P., Reichenbacher, L., Sauter, D., Allen, R., Scott, S., & Hill, E. (2012). Measuring the effects of alexithymia on perception of emotional vocalizations in autistic spectrum disorder and typical development. *Psychological Medicine, 42,* 2453-2459. doi:10.1017/S0033291712000621

Hedges, L. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review, 24*, 189-233. doi:10.1016/j.dr.2004.01.001

*Hollocks, M. J., Lerh, J. W., Magiati, I., Meiser-Stedman, R., & Brugha, T. S. (2018). Anxiety and depression in adults with autism spectrum disorder: A systematic review and meta-analysis. *Psychological Medicine,* 1-14. doi: 10.1017/S0033291718002283

Howell, R. T. & Shields, A. L. (2008). The file drawer problem in reliability generalization: A strategy to compute a fail-safe N with reliability coefficients. *Educational and Psychological Measurement, 68*(1), 120-128. doi: 10.1177/0013164407301528

Huang, C. (2017). Cross-informant agreement on the Child Behavior Checklist for youths: A meta-analysis. *Psychological Reports, 120,* 1096-1116. doi:10.1177/0033294117717733

Huberty, T. J. (2014). Best practices in school-based interventions for anxiety and depression. In P. L. Harrison & A. Thomas (Eds.), *Best Practices in School Psychology: Student-Level Services* (pp. 349-363). Bethesda, MD: National Association of School Psychologists.

Hudson, C. C., Hall, L., & Harnkess, K. L. Prevalence of depressive disorders in individuals with autism spectrum disorder: A meta-analysis. *Journal of Abnormal Child Psychology.* doi: 10.1007/s10802-018-0402-1

Humrichouse, J., Chmielewski, M., McDade-Montez, E. A., & Watson, D. (2007). Affect assessment through self-report methods. In J. Rottenberg & S. L. Johnson (Eds.), *Emotion and Psychopathology: Bridging Affective and Clinical Science* (pp. 13-34). Washington, DC: American Psychological Association.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

*Hurtig, T., Kuusikko, S., Mattila, M.-L., Haapsamo, H., Ebeling, H., Jussila, K., … Moilanen, I. (2009). Multi-informant reports of psychiatric symptoms among high-functioning adolescents with Asperger syndrome or autism. *Autism*, *13*(6), 583–598. doi: 10.1177/1362361309335719

*Jamison, T. R., & Schuttler, J. O. (2015). Examining social competence, self-perception, quality of life, and internalizing and externalizing symptoms in adolescent females with and without autism spectrum disorder: A quantitative design including between-groups and correlational analyses Understanding the links between sex/gender and autism Dr Meng-Chuan Lai. *Molecular Autism, 6*(1). doi: 10.1186/s13229-015-0044-x

Jang, J., Matson, J., Williams, L., Tureck, K., Goldin, R., & Cervantes, P. (2013). Rates of comorbid symptoms in children with ASD, ADHD, and comorbid ASD and ADHD. *Research in Developmental Disorders, 34*, 2369-2378. doi:10.1016/j.ridd.2013.04.021

Jensen, P. S., Traylor, J., Xenakis, S. N., & Davis, H. (1988). Child psychopathology rating scales and interrater agreement: I. Parents' gender and psychiatric symptoms. Journal of the American Academy of Child & Adolescent Psychiatry, 27, 442–450.

*Jepsen, M., Gray, K. M., & Taffe, J. R. (2012). Agreement in multi-informant assessment of behavior and emotional problems and social functioning in adolescents with autistic and asperger's disorder. *Research in Autism Spectrum Disorders, 6,* 1091-1098. doi: 10.1016/j.rasd.2012.02.008

*Joyce, C., Honey, E., Leekam, S. R., Barrett, S. L., & Rodgers, J. (2017). Anxiety, intolerance of uncertainty and restricted and repetitive behaviour: Insights directly from young people with ASD. *Journal of Autism and Developmental Disorders*, *47*(12), 3789–3802. doi: 10.1007/s10803-017-3027-2

*Kaat, A. J. (2014). Reliability of anxiety symptoms in youth with autism spectrum disorder (The Ohio State University). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/1798732246?accountid=12598

*Kaat, A. J., & Lecavalier, L. (2015). Reliability and validity of parent- and child-rated anxiety measures in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 45*, 3219-3231. doi:10.1007/s10803-015-2481-y

*Kanne, S. M., Abbacchi, A. M., Consttantino, J. N. (2009). Multi-informant ratings of psychiatric symptom severity in children with autism spectrum disorders: The importance of environmental context. *Journal of Autism and Developmental Disorders, 39,* 856-864. doi: 10.1007/s10803-009-0694-7

Kaur, M., Srinivasan, S. M., & Bhat, A. N. (2018). Comparing motor performance, praxis, coordination, and interpersonal synchrony between children with and without autism spectrum disorder (ASD). *Research in Developmental Disabilities, 72,* 79-95. doi: 10.1016/j.ridd.2017.10.025

*Keith, J. M., Jamieson, J. P., & Bennetto, L. (2018). The importance of adolescent self-report in autism spectrum disorder: Integration of questionnaire and autonomic measures. *Journal of*

Abnormal Child Psychology, 47(4), 741–754. doi: 10.1007/s10802-018-0455-1

Kent, R., & Simonoff, E. (2017). Prevalence of anxiety in autism spectrum disorders. In C. M. Kerns, P. Renno, E. A. Storch, P. C. Kendall, & J. J. Wood (Eds.), *Anxiety in children and adolescents with autism spectrum disorder: Evidence-based assessment and treatment* (pp. 5-32). San Diego, CA: Academic Press.

Kerns, C. M., & Kendall, P. C. (2014). Autism and anxiety: Overlap, similarities, and differences. In T. E. Davis III, S. W. White, & T. H. Ollendick (Eds.), *Handbook of autism and anxiety* (pp. 75-90). New York, NY: Springer International Publishing.

Kiep, M., & Spek, A. A. (2017). Executive functioning in men and women with an autism spectrum disorder. *Autism Research, 10,* 940-948. doi: 10.1002/aur.1721

Kim, Y. S., Leventhal, B. L., Koh, Y., Fombonne, E., Laska, E., Lim, E.,…Grinker, R. R. (2011). Prevalence of Autism spectrum disorders in a total population sample. *The American Journal of Psychiatry, 168,* 904-912. Retrieved from http://ajp.psychiatryonline.org

Klien, D. N., Dougherty, L. R., & Olino, T. M. (2005). Toward guidelines for evidence-based assessment of depression in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34* (3), 412-432. doi: 10.1207/s15374424jccp3403_3

Kovacs M. (2003). *Children's Depression Inventory.* North Tonawanda, NY: Multi-Health Systems Inc.

*Lane, B. R., Paynter, J., & Sharman, R. (2013). Parent and teacher ratings of adaptive and challenging behaviours in young children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 7*(10), 1196–1203. doi: 10.1016/j.rasd.2013.07.011

Layne, A. E., Bernstein, G. A., & March, J. S. (2009). Teacher awareness of anxiety symptoms in children. *Child Psychiatry and Human Development, 36*(4), 383-392. doi: 10.1007/s10578-006-0009-6

Lecavalier, L. (2014). Phenotypic variability in autism spectrum disorder: Clinical considerations. In T. E. Davis III, S. W. White, & T. H. Ollendick (Eds.), *Handbook of autism and anxiety* (pp. 15-30). New York, NY: Springer International Publishing.

*Lee, K. (2009). Predictors of depression in children with high -functioning autism spectrum disorders: The relationship between self -perceived social competence, intellectual ability, and depressive symptomology (State University of New York at Buffalo). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/305096160?accountid=12598

Leyfer, O. T., Folstein, S. E., Bacalman, S., Davis, N. O., Dinh, E., Morgan, J., et al. (2006). Comorbid psychiatric disorders in children with autism: Interview development and rates

of disorders. Journal of Autism and Developmental Disorders, 36(7), 849–861. doi:10.1007/s10803-006-0123-0.

Liu, J., Tang, W., Chen, G., Lu, Y., Feng, C., & Tu, X. M. (2016). Correlation and agreement: Overview and clarification of competing concepts and measures. *Shanghai Archives of Psychiatry, 28,* 115-120.

*Lohr, W. D., Daniels, K., Wiemken, T., Williams, P. G., Kelley, R. R., Kuravackel, G., & Sears, L. (2017). The screen for child anxiety-related emotional disorders is sensitive but not specific in identifying anxiety in children with high-functioning autism spectrum disorder: A pilot comparison to the achenbach system of empirically based assessment scales. *Frontiers in Psychiatry, 8,* 10. doi: 10.3389/fpsyt.2017.00138

*Lopata, C., Toomey, J., Fox, J. D., Volker, M. A., Chow, S. Y., Thomeer, M. L.,…Smerbeck, A. M. (2010). Anxiety and depression in children with HFSADs: Symptom levels and source differences. *Journal of Abnormal Child Psychology, 38,* 765-776. doi:10.1007/s10802-010-9406-1

*Luxford, S., Hadwin, J. A., & Kovshoff, H. (2017). Evaluating the effectiveness of a school-based cognitive behavioural therapy intervention for anxiety in adolescents diagnosed with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 47*(12), 3896–3908. doi: 10.1007/s10803-016-2857-7

MacDonald, M., Lord, C., Ulrich, D. A. (2014). Motor skills and calibrated autism severity in young children with autism spectrum disorder. *Adapted Physical Activity Quarterly, 31,* 95-105. doi: 10.1123/apaq.2013-0068

*Magiati, I., Chan, J. Y., Tan, W. J., Poon, K. K. (2014). Do non-referred young people with autism spectrum disorders and their caregivers agree when reporting anxiety symptoms? A preliminary investigation using the Spence Children's Anxiety Scale. *Research in Autism Spectrum Disorders, 8,* 546-558. doi: 10.1016/j.rasd.2014.01.015

Magiati, I., Ozsivadjian, A., Kerns, C. M. (2017). Phenomenology and presentation of anxiety in autism spectrum disorder In C. M. Kerns, P. Renno, E. A. Storch, P. C. Kendall, & J. J. Wood (Eds.), *Anxiety in children and adolescents with autism spectrum disorder:*

Evidence-based assessment and treatment (pp. 33-54). San Diego, CA: Academic Press.

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,* 1304-1313.

March, J. A. (1997). *Multidimensional Anxiety Scale for Children*. North Tonawanda, NY: Multi-Health Systems.

March, J. S., James, P. D., Sullivan, K., Stallings, P., & Conners, K. (1997). The Multidimensional Anxiety Scale for Children (MASC) factory structure, reliability, and validity. *The Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 554-565. doi:10.1097/00004583-199704000-00019

March, J. S. & MHS Staff. (1997). *Multidimensional Anxiety Scale for Children Technical Manual.* North Tonawanda, NY: Multi-Health Systems.

Matson, J. L., & Williams, L. W. (2014). Depression and mood disorders among persons with autism spectrum disorders. *Research in Developmental Disabilities, 35*, 2003-2007. doi:10.1016/j.ridd.2014.04.020

Mayo, J., Chelbowski, C., Fein, D. A., & Eigsti, I. (2013). Age of first words predicts cognitive ability and adaptive skills in children with ASD. *Journal of Autism and Developmental Disorders, 43,* 253-264. doi: 10.1007/s10803-012-1558-0

Mazefsky, C. A., Kao, J., & Oswald, D. P. (2011). Preliminary evidence suggesting caution in the use of psychiatric self-report measures with adolescents with high-functioning autism spectrum disorders. *Research in Autism Spectrum Disorder, 5,* 164-174.

Mazzone, L., Ruta, L., & Reale, L. (2012). Psychiatric comorbidities in asperger syndrome and high functioning autism: Diagnostic challenges. *Annals of General Psychiatry, 11*(16). 1-13. Retrieved from http://www.annals-general-psychiatry.com/content/11/1/16

*McDonald, C. A., Lopata, C., Donnelly, J. P., Thomeer, M. L., Rodgers, J. D., & Jordan, A. K. (2016). Informant discrepancies in externalizing and internalizing symptoms and adaptive skills of high-functioning children with autism spectrum disorder. *School Psychology Quarterly, 31*(4), 467-477. doi: 10.1037/spq0000150

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46.

McPhillips, M., Finlay, J. Bejerot, S., & Hanley, M. (2014). Motor deficits in children with autism spectrum disorder: A cross-syndrome study. *Autism Research, 7*, 664-676. doi: 10.1002/aur.1408

Merrell, K. W. (2008). Understanding internalizing problems in Helping Students Overcome Depression and Anxiety: A Practical Guide. New York, NY: Guilford Press

Merrell, K., McClun, L. A., Kempf, K. G., & Lund, J. (2002). Using self-report assessment to identify children with internalizing problems: Validity of the internalizing symptoms scale for children. *Journal of Psychoeducational Assessment, 20,* 223-239.

*Mertens, J., Zane, E. R., Neumeyer, K., & Grossman, R. B. (2017). How anxious do you think I am? Relationship between state and trait anxiety in children with and without ASD during social tasks. *Journal of Autism and Developmental Disorders*, *47*(12), 3692–3703. doi:

10.1007/s10803-016-2979-y

Michael, K. D., & Merrell, K. W. (1998). The reliability of children's self-reported internalizing symptoms over brief- to medium-length time intervals. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*(2), 194-201.

Minshawi, N, F. (2008). Behavioral assessment and treatment of self-injurious behavior in autism. *Child and Adolescent Psychiatric Clinics of North America, 17,* 875-886. doi: 10.1016/j.chc.2008.06.012

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med, 6, e1000097. doi:10.1371/journal.pmed1000097

Nardi, D. A. (2007). Depression assessment & early intervention. *Journal of Psychosocial Nursing, 45*(3), 48-51.

Nebel-Schwalm, M., & Worley, J. (2014). Other disorders frequently comorbid with autism. In T. E. Davis III, S. W. White, & T. H. Ollendick (Eds.), *Handbook of autism and anxiety* (pp. 47-60). New York, NY: Springer International Publishing.

*Neil, L., White, H., Warren, K., & Pellicano, E. (2019). Anxiety and Interpretation of Ambiguity in Autistic Children, Typical Children and Their Mothers. *Journal of Autism and Developmental Disorders*, *49*(3), 1035–1047. doi: 10.1007/s10803-018-3781-9

*Ooi, Y P, Lam, C. M., Sung, M., Tan, W. T. S., Goh, T. J., Fung, D. S. S., … Chua, A. (2008). Effects of cognitive-behavioural therapy on anxiety for children with high-functioning autistic spectrum disorders. *Singapore Medical Journal*, *49*(3), 215–220.

*Ooi, Yoon Phaik, Weng, S.-J., Magiati, I., Ang, R. P., Goh, T. J., Fung, D. S., & Sung, M. (2016). Factors influencing agreement between parent and child reports of anxiety symptoms among children with high-functioning Autism Spectrum Disorders. *Journal of Developmental and Physical Disabilities*, *28*(3), 407–424. doi: 10.1007/s10882-016-9481-5

Ozinci, Z., Kahn, T., & Antar, L. N. (2012). Depression in patients with autism spectrum disorder. *Psychiatric Annals, 42*(8), 293-295. doi: 10.3928/00485713-20120806-06

*Ozsivadjian, A., Hibberd, C., & Hollocks, M. J. (2014). Brief report: The use of self-report measures in young people with autism spectrum disorder to access symptoms of anxiety, depression and negative thoughts. *Journal of Autism and Developmental Disorders, 44*, 969-974. doi:10.1007/s10803-013-1937-1

Park, S., Park, M., Kim, H. J., & Yoo, H. J. (2013). Anxiety and depression symptoms in children with Asperger syndrome compared with attention-deficit/hyperactivity disorder and depressive disorder. *Journal of Child and Family Studies, 22*, 559-568. doi: 10.1007/s10826-012-9611-3

Pautasso, M. (2010). Worsening file-drawer problem in the abstracts of natural, medical, and social science databases. *Scientometrics, 85*(1), 193-202. doi: 10.1007/s11192-010-0233-5

Pellicano, E. (2012). The development of executive function in autism. *Autism Research and Treatment, 2012*, 1-8. doi:10.1155/2012/146132

*Peterson, J. L. (2017). Internalizing symptoms: Relations to executive functions in young children with autism spectrum disorder (Seattle Pacific University). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/1823285519?accoun acc=12598

*Pisula, E., Pudło, M., Słowińska, M., Kawa, R., Strząska, M., Banasiak, A., & Wolańczyk, T. (2017). Behavioral and emotional problems in high-functioning girls and boys with autism spectrum disorders: Parents' reports and adolescents' self-reports. *Autism, 21*(6), 738–748. doi: 10.1177/1362361316675119

Pitzianti, M., D'Agati, E., Pontis, M., Baratta, A., Casarelli, L., Spiridigliozzi, S.,…Pasini, A. (2016). *Journal of Psychiatric Practice, 22*, 22-30. doi:10.1097/PRS.0000000000000120

Ratz, C., & Lenhard, W. (2013). Reading skills among students with intellectual disabilities. *Research in developmental disabilities*, *34*(5), 1740–1748. https://doi.org/10.1016/j.ridd.2013.01.021

*Reaven, J. A., Blakeley-Smith, A., Nichols, S., Dasari, M., Flanigan, E., & Hepburn, S. (2009). Cognitive-behavioral group treatment for anxiety symptoms in children with high-functioning autism spectrum disorders: A pilot study. *Focus on Autism and Other Developmental Disabilities*, *24*(1), 27–37. doi: 10.1177/1088357608327666

Rebekić, A., Lončarić, Z., Petrović, S., & Marić, S. (2015). Pearson's or spearman's correlation coefficient—Which one to use? *Poljoprivreda (Agriculture)*, *21*(2), 47-54. doi: 10.18047/poljo.21.2.8

Reynolds, C. R. & Kamphaus, R. W. (2004). *Behavior Assessment System for Children* (2nd ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Rice, C. (2007). Prevalence of Autism Spectrum Disorders—Autism and developmental disabilities monitoring network, six sites, United States, 2000. *Autism and Developmental Disabilities Monitoring*, 1-11

*Richdale, A. L., & Baglin, C. L. (2015). Self-report and caregiver-report of sleep and psychopathology in children with high-functioning autism spectrum disorder: A pilot study. *Developmental Neurorehabilitation*, *18*(4), 272–279. doi: 10.3109/17518423.2013.829534

Richters, J. E. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. Psychological Bulletin, 112, 485–499.

*Rodgers, J., Wigham, S., McConachie, H., Freeston, M., Honey, E., & Parr, J. R. (2016). Development of the anxiety scale for children with autism spectrum disorder (ASC-ASD). *Autism Research*, *9*(11), 1205–1215. doi: 10.1002/aur.1603

*Rodriguez, G. R. (2017). Examining the Relations between Psychopathology and Social Skills in Children with ASD. Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/2011263266?accountid=12598

*Rosen, T. E., & Lerner, M. D. (2018). Error-related brain activity and anxiety symptoms in youth with autism spectrum disorder. *Autism Research*, *11*(2), 342–354. doi: 10.1002/aur.1898

*Rosen, T. E., Spaulding, C. J., Gates, J. A., & Lerner, M. D. (2019). Autism severity, co-occurring psychopathology, and intellectual functioning predict supportive school services for youth with autism spectrum disorder. *Autism : The International Journal of Research and Practice*, 1362361318809690. doi: 10.1177/1362361318809690

*Rosenberg, R. (2016). The influence of social skills on the development of internalizing disorders in children with autism spectrum disorder (Widener University). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/1694871733?accountid=12598

*Rump, K. M. (2012). Affective experiences in adolescents with autism: An EMA study (University of Pittsburgh). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/1025851008?accountid=12598

*Schiltz, H. K., McVey, A. J., Barrington, A., Haendel, A. D., Dolan, B. K., Willar, K. S., … Van Hecke, A. V. (2018). Behavioral inhibition and activation as a modifier process in autism spectrum disorder: Examination of self-reported bis/bas and alpha eeg asymmetry. *Autism Research*. doi: 10.1002/aur.2016

*Schwartz, C. B. (2010). Predicting variation in social outcome among adolescents with high-functioning autism. Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/622199280?accountid=12598

*Sharpley, C. F., Bitsika, V., Agnew, L. L., & Andronicos, N. M. (2015). Eight-month test–retest agreement in morning salivary cortisol, self- and parent-rated anxiety in boys with an Autism Spectrum Disorder. *Physiology & Behavior*, *151*, 207–212. doi: 10.1016/j.physbeh.2015.07.027

Silverman, W. K., & Ollendick, T. H. (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child and Adolescent*

Psychology, 34(3), 380-411.

*Slavin, M. J. (2010). Evaluating the effectiveness of a multidimensional social skills intervention program for middle school students with Autism Spectrum Disorders (Fairleigh Dickinson University). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/843220680?accountid=12598

Spek, A. A., Scholte, E. M., & Can Berckalaer-Onnes, I. A. (2010). Theory of mind in adults with HFA and Asperger syndrome. *Journal of Autism and Developmental Disorders, 40,* 280-289. doi:10.1017/s10803-009-0860-y

Spence, S. H. (1998). A measure of anxiety symptoms among children. *Behavior Research and Therapy, 36,* 545-566.

*Sterling, L., Renno, P., Storch, E. A., Ehrenreich-May, J., Lewin, A. B., Arnold, E., … Wood, J. (2015). Validity of the Revised Children's Anxiety and Depression Scale for youth with autism spectrum disorders. *Autism*, *19*(1), 113–117. doi: 10.1177/1362361313510066

*Stern, J. A., Gadgil, M. S., Blakeley-Smith, A., Reaven, J. A., & Hepburn, S. L. (2014). Psychometric properties of the SCARED in youth with Autism spectrum disorder. *Research in Autism Spectrum Disorders*, *8*(9), 1225–1234. doi: 10.1016/j.rasd.2014.06.008

*Storch, E. A., Lewin, A. B., Collier, A. B., Arnold, E., De Nadai, A. S., Dane, B. F., … Murphy, T. K. (2015). A randomized controlled trial of cognitive-behavioral therapy versus treatment as usual for adolescents with autism spectrum disorders and comorbid anxiety. *Depression and Anxiety*, *32*(3), 174–181. doi: 10.1002/da.22332

*Stratis, E. A., & Lecavalier, L. (2017). Predictors of Parent–Teacher Agreement in Youth with Autism Spectrum Disorder and Their Typically Developing Siblings. *Journal of Autism and Developmental Disorders*, *47*(8), 2575–2585. doi: 10.1007/s10803-017-3173-6

Stratis, E. A., & Lecavalier, L. (2015). Informant agreement for youth with autism spectrum disorder or intellectual disability: A meta-analysis. *Journal of Autism and Developmental Disorders, 45,* 1026-1041. doi: 10.1007/s10803-014-2258-8

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research, 6*(6). doi: 10.1002/aur.1329

*Taylor, J. M., Volker, M. A., Rispoli, K. M, Rodgers, J. D., Thomeer, M., Lopata, C.,… Smerbeck, A. (2018). Depression, anxiety, and hyperactivity in youth with HFASD: A replication and extension of symptom level differences in self-report vs. parent report. *Journal of Autism and Developmental Disorders.* doi*:*10.1007/s10803-018-3779-3

Theoharides, T. C. & Zhang, B. (2011). Neuro-inflammation, blood-brain barrier, seizures and

autism. *Neuroinflammation, 8,* 168-173. doi: http://www.jneuroinflammation.com/content/8/1/168

Tuchman, R. (2013). Autism and seizures. *Epilepsy & Behavior, 28*(2), 311. doi: 10.1016/j.yebeh.2012.04.033

*Ung, D., Boone, D. M., McBride, N., Howie, F., Scalli, L., & Storch, E. A. (2017). Parent and teacher agreement of behavior problems in youth diagnosed with and without autism spectrum disorders. *Journal of Child and Family Studies*, *26*(2), 370–380. doi: 10.1007/s10826-016-0566-7

Ung, D., Wood, J. J., Ehrenreich-May, J., Arnold, E. B., Fujii, C., Renno, P., . . . Storch, E. A. (2013). Clinical characteristics of high-functioning youth with autism spectrum disorder and anxiety. *Neuropsychiatry, 3,* 147-157. doi:http://dx.doi.org.proxy1.cl.msu.edu/10.2217/npy.13.9

*van Schalkwyk, G., Smith, I. C., Silverman, W. K., & Volkmar, F. R. (2018). Brief report: Bullying and anxiety in high-functioning adolescents with ASD. *Journal of Autism and Developmental Disorders*, *48*(5), 1819–1824. doi: 10.1007/s10803-017-3378-8

van Steensel, F. J. A., & Heeman, E. J. (2017). Anxiety levels in children with autism spectrum disorder: A meta-analysis. *Journal of Child and Family Studies, 26,* 1753-1767. doi:10.1007/s10826-017-0687-7.

Van Wigngaarden-Cremers, P. J., van Eeten, E., Groen, W. B., van Deurzen, P. A., Oosterling, I. J., & van der Gaag, R. J. (2014). Gender and age differences in the core triad of impairments in autism spectrum disorders: a systematic review and meta-analysis. *Journal of Autism and Developmental Disorders, 44,* 627-635.

Vasa, R. A., Mazurek, M. O., Mahajan, R., Bennett, A. E., Pilar Bernal, M., Nozzolillo, A. A.,…& Coury, D. L. (2016). Assessment and treatment of anxiety in youth with autism spectrum disorder. *Pediatrics, 137,* 116-123. doi: 10.1542/peds.2015-2851J

*Vickerstaff, S., Heriot, S., Wong, M., Lopes, A., & Dossetor, D. (2007). Intellectual ability, self-perceived social competence, and depressive symptomatology in children with high-functioning autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*(9), 1647–1664. doi: 10.1007/s10803-006-0292-x

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3). Retrieved from http://www.jstatsoft.org/

Volker, M. A., (2012). Introduction to the special issue: High-functioning autism spectrum disorders in the schools. *Psychology in the Schools, 49,* 911-916. doi:10.1002/pits.21653

Volker, M. A. & Lopata, C. (2008) Autism: A review of biological bases, assessment, and intervention. *School Psychology Quarterly, 23*(2), 258-270.

Volker, M. A., Lopata, C., Smerbeck, A. M., Knoll, V. A., Thomeer, M. L., Toomey, J. A., & Rodgers, J. D. (2010). BASC-2 PRS profiles for students with high functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders, 40*, 188-199. doi:10.1007/s10803-009-0849-6

Whitcomb, S. A., & Merrell, K. W. (2013). *Behavioral, social, and emotional assessment of children and adolescents* (4th ed.). New York, NY: Routledge.

White, S. W., Schry, A. R., & Maddox, B. B. (2012). Brief report: The assessment of anxiety in high-functioning adolescents with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 42*, 1138–1145.

*Whitehead, J. L. (2005). Treating AD -related anxiety as measured by the BASC in adolescents with Asperger's disorder (Argosy University/Seattle). Retrieved from http://ezproxy.msu.edu/login?url=https://search.proquest.com/docview/305358075?accountid=12598

*Wijnhoven, L. A. M. W., Creemers, D. H. M., Vermulst, A. A., & Granic, I. (2018). Prevalence and Risk Factors of Anxiety in a Clinical Dutch Sample of Children with an Autism Spectrum Disorder. *Frontiers in Psychiatry*, *9*, 50. doi: 10.3389/fpsyt.2018.00050

*Wood, J. J., Drahota, A., Sze, K., Har, K., Chiu, A., & Langer, D. A. (2009). Cognitive Behavioral Therapy for Anxiety in Children with Autism Spectrum Disorders: A Randomized, Controlled Trial. *Journal of Child Psychology and Psychiatry, 50*(3), 224–234. doi: 10.1111/j.1469-7610.2008.01948.x

*Wood, J. J., Ehrenreich-May, J., Alessandri, M., Fujii, C., Renno, P., Laugeson, E., … Storch, E. A. (2015). Cognitive behavioral therapy for early adolescents with autism spectrum disorders and clinical anxiety: A randomized, controlled trial. *Behavior Therapy, 46*(1), 7–19. doi: 10.1016/j.beth.2014.01.002