Predicting Growth in Prosocial and Externalizing Behaviors During a Summer Treatment Program for Children with ASD, ADHD, and Comorbid ASD and ADHD

Shannon Ludeman

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Kristen Missall, Chair

Ilene Schwartz

Elizabeth Sanders

Benjamin Aaronson

Program Authorized to Offer Degree:

College of Education

University of Washington

Abstract

Predicting Growth in Prosocial and Externalizing Behaviors During a Summer Treatment Program for Children with ASD, ADHD, and Comorbid ASD and ADHD

Shannon Ludeman

Chair of the Supervisory Committee:

Kristen Missall

College of Education

Children ages 6-12 with Autism Spectrum Disorders (ASD) and Attention-Deficit/Hyperactivity Disorder (ADHD) often demonstrate similar challenges in in social skills and externalizing behavior (e.g., Leitner, 2014; May et al., 2018). Although distinct evidence-based intervention strategies for improving social skills and externalizing behavior have been shown to have positive effects for children with ASD (e.g., Wong et al., 2015) as well as children with ADHD (e.g., Fabiano et al., 2014), less is known about how a single evidence-based intervention might differentially affect children with ASD, ADHD, or both (Davis & Kollins, 2012).This study examined behavioral trajectories of children with ASD (n = 39), ADHD (n = 50), or both (n = 28) who participated in Apex Summer Camp prior to the COVID-19 pandemic. Apex is an intensive behavioral intervention modeled after the Summer Treatment Program (Pelham et al.,

2012). Hierarchical linear modeling was used to model growth in prosocial and externalizing behaviors during camp, and to test whether diagnosis and pre-camp parent-reported child behavior ratings (i.e., Social Responsiveness Scale and Child Behavior Checklist) predicted growth. Results showed that all children, on average, had significant increases in most prosocial behaviors, but no overall changes in most externalizing behaviors. More importantly, campers with ASD and comorbid ASD and ADHD were rarely different from their counterparts with ADHD alone, though comorbid diagnosis was more often associated with less desirable outcomes. Last, parent pre-intervention behavior ratings were largely not predictive of positive behavior growth but predicted the intercept of some externalizing behaviors. Implications for clinical and school-based practice are discussed.

Keywords: ASD, ADHD, social skills, externalizing behavior, parent rating scales

Acknowledgements

I extend my deepest fathomable gratitude to Apex staff and families, both past and present, for playing a role in my growth as a clinician and researcher. Thank you to Akshay Venkatesh who assisted with the early stages of cleaning this seemingly indomitable dataset. Thank you to Will McCloud, Dr. Alexa Matlack, Dr. Zoey Phillips, Caitlin Courshon, Taylor McNamara, Sam Peoples, Dr. Tyler Sasser, and the many other members of the leadership team for your inspiring examples and friendship. And, of course, thank you to Dr. Ben Aaronson for continuing to support my professional and personal growth and your tireless dedication to the staff and families of Apex. I am completely certain that I would not be here without you.

I must also acknowledge the invaluable contributions of my committee to this work and to my education. Thank you to Dr. Spieker who generously supported my involvement as a LEND trainee, Dr. Schwartz who has long been a supporter of Apex and touched my training in many different capacities, and Dr. Sanders who taught me almost everything I know about statistics and fielded my many questions throughout this process. Thank you to my advisor, Dr. Missall; I have grown immeasurably as a writer and researcher thanks to you. Thank you also to my colleagues and faculty in the school psychology program for showing me what this field can be.

Lastly, I thank my friends and family for their unending support, and Alex for being not just my resident Excel expert but also my loving partner in all things. I am forever in awe of my wonderful support system.

"It's a lesson I have learned over and over again, but it bears repeating: no one achieves anything alone." - Leslie Knope, *Parks and Recreation*

Table of Contents

Chapter 1: Introduction

Attention deficit hyperactivity disorder (ADHD) and autism spectrum disorders (ASD) are neurodevelopmental disorders that tend to surface in early or middle childhood (American Psychiatric Association [APA], 2013). ADHD is most often marked by persistent inattention and overactivity (APA, 2013). ASD is identified by pervasive difficulties in social communication and restricted interests or repetitive behaviors (APA, 2013). In many cases, manifestation of these disorders shares some degree of symptomatology, such as children with ASD demonstrating problems with executive function or children with ADHD struggling to maintain peer relationships (e.g., Leitner, 2014). Although externalizing behavior is not a core symptom of either ADHD or ASD, it often accompanies both diagnoses (e.g., APA, 2013; Craig et al., 2016). ADHD and ASD also tend to be associated with a variety of other differences in development, including gastrointestinal (Buie et al., 2010; McKeown et al., 2013), sleep (Lindor et al., 2019; Scott et al., 2013), motor (APA, 2013), and sensory problems (Bijlenga et al., 2017). The degree of overlap in symptomatology and shared features is so great that comorbidity of ADHD and ASD is estimated to occur in up to 83% of cases (May et al., 2018), and it is often difficult for clinicians and researchers to distinguish between them (Grzadzinski et al., 2016).

Challenges with Diagnosis of ADHD and ASD

The conceptual relationship between ADHD and ASD is poorly defined in the published literature. Competing theories range from viewing ADHD and ASD as one large continuum (Kern et al., 2012; van der Meer et al., 2012) to viewing them as wholly distinct disorders (Green et al., 2016; Rommelse et al., 2018). It is perhaps unsurprising, then, that recommended diagnostic procedures often fail to clearly differentiate between these diagnoses. Research has

suggested that many common measures used in assessment and treatment contexts have poor discriminant validity between ADHD and ASD (Grzadzinski et al., 2016; Yerys et al., 2016). For example, some authors have found that a substantial portion of children with ADHD alone met criteria for ASD on the Social Responsiveness Scale, Second Edition (SRS-2; Constantino & Gruber, 2012; Joshi et al., 2011; Reiersen et al., 2007), a common measure of social skills that was designed for use with children with ASD. Other common assessment tools such as the ADHD Rating Scale-IV (DuPaul et al., 1998), the Autism Diagnostic Interview, Revised (ADI-R; Rutter, LeCouteur, et al., 2003), and the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord & Rutter, 2012) have evidence of limited discriminant validity between ADHD and ASD (Grzadzinski et al., 2016; Yerys et al., 2016). Many families have received a diagnosis of ASD years after a diagnosis of ADHD, delaying the onset of critical early intervention services (Miodovnik et al., 2015). Despite these challenges with determining the appropriate diagnostic category, clinicians and researchers often use diagnosis to determine which families receive certain services and participate in certain kinds of research (American Academy of Child and Adolescent Psychiatry [AACAP], 2007; Volkmar et al., 2014). The implication of relying on a complicated and imperfect diagnostic process to determine the course of psychosocial treatment for children with ADHD and children with ASD is that many children may not receive the full complement of available interventions to support their social and behavioral development. In other words, rather than relying on a diagnosis to determine the course of treatment in cases of suspected ADHD or ASD, clinicians and researchers may find it more useful to determine a child's unique areas of need and recommend interventions accordingly.

Commonalities in Recommended Intervention Practices for ADHD and ASD

In addition to overlap in clinical presentation, there are also striking similarities among evidence-based intervention strategies and programs recommended for children with ADHD and children with ASD. Examination of these respective literatures reveals a number of well-supported intervention strategies and principles that are shared, such as foundations in behaviorism, specifically applied behavioral analysis (Fabiano et al., 2014; Wong et al., 2015). In addition, the goals of intervention are often similar and tend to focus on social skills, externalizing behavior, and generalization of skills to new environments (Chronis et al., 2006; Pelham & Fabiano, 2008; Pelham et al., 2012; Wong et al., 2015). Interventions incorporate principles of contingency management, involve caregivers, provide opportunities for practice and coaching in naturalistic environments, and are most effective at a high intensity (i.e., interventions are administered for a significant number of hours each week; Fabiano et al., 2014; National Research Council [NRC], 2001).

Parent involvement in intervention is one aspect of most evidence-based interventions for both children with ADHD and children with ASD that warrants special attention. Gathering information from parents is a necessary part of the diagnostic and intervention process as parents play a critical role in supporting children with ADHD and children with ASD throughout their lifespan (e.g., AACAP, 2007; Williams White et al., 2007). Providing opportunities for parents to engage in interventions by receiving training or completing measures that better inform treatment decisions can enhance the value of psychosocial interventions for children with ADHD (e.g., Reed et al., 2017) and children with ASD (e.g., Wong et al., 2015). For this reason, multimethod assessment practices, such as combinations of rating scales and direct observations, are often

recommended when evaluating outcomes associated with interventions among these populations (Barkley, 2015; Hyman et al., 2020).

Given both the similarities in impacted areas of development and evidence for similar intervention practices across ADHD and ASD, it might be expected that these populations would be frequently combined for the purposes of intervention. However, published intervention research that incorporates both children who have ADHD and children who have ASD is rare (Davis & Kollins, 2012; Reirsen & Todd, 2008), as is research to support intervention strategies for children with comorbid diagnoses (Antshel et al., 2016). Some research has suggested that many children with comorbid ADHD and ASD are not treated for both disorders, in part due to the limited knowledge about how best to support children with comorbid ADHD and ASD (Joshi et al., 2017). One intervention, the Summer Treatment Program (STP; Pelham et al., 2012), is a collection of intervention strategies that are empirically supported for children ages 6-12 with ADHD. Participation in the STP has been associated with positive outcomes for peer relationships, on-task behavior, and challenging behaviors (Chronis et al., 2004; Pelham & Fabiano, 2008; Pelham et al., 2000), as well as increased knowledge of sports rules and skills and improved sportsmanship (O'Connor et al., 2013). The STP's major components have many striking similarities with empirically-validated interventions for children with ASD.

Limitations of Existing Literature

Despite commonalities in evidence-based research, there are only three peer-reviewed studies of the STP to date that included children with ASD. Two of these studies were single-case designs, and all participants were boys with what authors described as "high functioning" autism. A small number of group design studies of the STP have attempted to model change over time by using behavioral data (Dunne, 2020; Mikami et al., 2010; Mitchell et al., 2015; Pelham et al.,

2000); only two of these studies accounted for membership in small intervention groups in their final models using hierarchical linear modeling (Dunne, 2020; Mikami et al., 2010). Furthermore, although previous research studies have included parent rating scales administered before and after implementing the STP (Mitchell et al., 2015; Pelham et al., 2000), no published studies available at the time of this review have analyzed the relationship between parent-report measures and behavioral data collected during the STP to determine the degree of consistency, meaning that there is little information available to inform the clinical utility of collecting parent rating scales prior to intensive behavioral interventions for children with ADHD and children with ASD. Taken together, this information points to two logical and crucial directions for future research: examination of social and behavioral outcomes by combining children with ADHD, children with ASD, and children with comorbid diagnoses for intervention; and the use of multimethod assessment strategies to predict and evaluate outcomes associated with psychosocial interventions for these children.

Purpose of the Current Study

The current study seeks to investigate these research directions by examining social and behavioral growth over the course of one intervention, Apex Summer Camp, which is a modified version of the Summer Treatment Program (STP; Pelham et al., 2012) that includes children with ADHD, children with ASD, and children with comorbid ADHD and ASD between the ages of 6 and 12. This study will examine the predictive value of diagnosis on social and behavioral change during camp and the potential impact of parent rating scales on this relationship. Findings from this study may support practitioners and researchers as they make decisions regarding which children make up intervention groups and research samples, as well has how to use multiple data sources to predict and evaluate social and behavioral outcomes. It may also

contribute to the field's emerging understanding of how to treat children with comorbid ADHD and ASD.

Chapter 2: Literature Review

Considerations for ADHD and ASD Diagnoses

The process of arriving at a diagnosis of ADHD or ASD has direct implications for treatment (AACAP, 2007; Volkmar et al., 2014), making examination of the diagnostic process a worthy exercise for researchers interested in investigating evidence-based interventions for children with ASD and children with ADHD. Both ADHD and ASD are neurodevelopmental disorders that tend to surface in early or middle childhood and are more prevalent than many other childhood disorders (APA, 2013). ADHD is the most diagnosed childhood mental illness, affecting approximately 11% of children in the United States (U.S.; National Institute of Mental Health [NIMH], 2017). ASD is also frequently diagnosed, currently estimated to impact approximately one in 54 eight-year-olds in the U.S. (Maenner et al., 2020). ADHD and ASD often impact individuals throughout their lifespan, resulting in long-reaching consequences for education, the workforce, and many other aspects of society (APA, 2013; Sasser et al., 2016). In a substantial number of cases, overlap in expressions of ADHD and ASD is high (e.g., Leitner, 2014). Both ADHD and ASD tend to manifest in some combination of executive dysfunction, externalizing behavior, and social problems (APA, 2013). Other problems, such as sensory challenges (APA, 2013; Bijlenga et al., 2017), disrupted sleep (Lindor et al., 2019; Scott et al., 2013), gastrointestinal problems (Buie et al., 2010; McKeown et al., 2013), and delayed fine and gross motor development (APA, 2013) are also common in cases of both ADHD and ASD.

Because of shared features related to core diagnostic criteria and common related problems, distinguishing between ADHD and ASD can be difficult and many children are eventually diagnosed with both; estimates of co-occurring ADHD and ASD range from 40% to 83% of cases (May et al., 2018). Children with comorbid diagnoses are likely to have more

severe presentations of both disorders as well as more impacted adaptive skills, leading to a greater degree of impairment and poorer developmental trajectory throughout the lifespan (e.g., Ashwood et al., 2015; Craig et al., 2015; Factor et al., 2017). Accurate and comprehensive assessment of domains and degree of impairment is tied directly to appropriate treatment decisions (e.g., American Academy of Child and Adolescent Psychiatry [AACAP], 2007; Volkmar et al., 2014); therefore, screening and diagnostic practices for ADHD and ASD are of great importance.

ADHD Symptoms and Diagnostic Process

ADHD is defined by executive dysfunction that occurs in the form of inattention, hyperactivity, or both (APA, 2013). Inattentive symptoms may be expressed through distractibility, careless mistakes, limited follow-through on tasks, poor organization, forgetfulness, and avoidance of tasks that require sustained cognitive effort, among others (APA, 2013). Hyperactive/impulsive symptoms often manifest through fidgeting, running or climbing more than other children, blurting out answers, excessive talking, or appearing as if "driven by a motor" (APA, 2013, p. 60). Several symptoms must be present before the age of 12 and appear in at least two settings, most often home and school (APA, 2013). Symptoms must interfere with social, academic, or occupational function and are not better explained by another disorder or substance (APA, 2013). Specifiers include subtype (i.e., predominantly inattentive, predominantly hyperactive/impulsive, or combined presentation), severity based on number of symptoms (i.e., mild, moderate, severe), and partial remission status (APA, 2013). Approximately one third of individuals diagnosed with ADHD in childhood retain a full diagnosis in adulthood (APA, 2013). Although social problems are not a core feature of this

diagnosis, many children with ADHD experience peer rejection and other social challenges (APA, 2013; Bagwell et al., 2001; Demopoulos et al., 2013).

Evaluations to determine the presence of ADHD usually begin with an interview to gather family medical and psychological history and determine the current impact of present ADHD symptoms (AACAP, 2007). Clinicians should gather information regarding developmental history; family constellation, dynamic, stressors, and strengths; and relevant concerns about the child's functioning from parents and teachers when possible (AACAP, 2007; Barkley, 2015). Initial parent and teacher interviews are often followed by rating scales (AACAP, 2007). Completion of both broad- (i.e., social-emotional development) and narrow-band (i.e., ADHD-specific) rating scales by both parents and teachers has evidenced strong sensitivity the presence of ADHD (Tripp et al., 2006), and has demonstrated robust diagnostic validity among children with ADHD compared to other methods of assessment (Pelham et al., 2005). Child characteristics such as limited language ability may result in lower agreement between parent and teacher rating scales (Gooch et al., 2017). Teacher-completed rating scales tend to out-perform parent-report in their relative accuracy, a finding that reinforces both the importance of collecting information from multiple informants throughout the assessment process and the relevance of the school setting to diagnostic decisions (Tripp et al., 2006).

Initial rating scales should include at least one broad-band measure, such as the Behavior Assessment System for Children, Third Edition (BASC-3; Reynolds & Kamphaus, 2015) or the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) to assess the child's overall functioning in addition to a measure specific to ADHD (AACAP, 2007; Barkley, 2015). Many such narrow-band scales have been validated for clinical use with children with ADHD, including the ADHD Rating Scale IV (DuPaul et al., 1998); Behavior Rating Inventory of

Executive Functioning, Second Edition (BRIEF-2; Gioia et al., 2015); Conners, Third Edition (Conners, 2008); Brown Executive Function/Attention Scales (Brown, 2018); and the Vanderbilt (National Institute for Children's Health Quality & American Academy of Pediatrics, 2002). To evaluate the severity of the core symptoms of ADHD, clinicians may ask parents and teachers to complete the Home Situations Questionnaire and School Situations Questionnaire (Barkley & Edelbrock, 1987), respectively (Barkley, 2015). Related measures such as the Weiss Functional Impairment Rating Scale (Canadian Attention Deficit Hyperactivity Disorder Resource Alliance, 2011) compliment other questionnaires by asking for information about the quality of life for children and adolescents with ADHD. It may also be of interest to evaluate the child's social skills using other narrow-band measures, such as the SRS-2 or the Social Skills Improvement System (Gresham & Elliot, 2008).

Often, agreement between raters is limited; therefore, additional assessment methods are recommended before making a clinical decision regarding an ADHD diagnosis (e.g., Sims & Lonigan, 2012). Clinicians should spend sufficient time with the child to develop a sense of their disposition and collect self-report measures, if the child's age permits (Barkley, 2015). Children with ADHD are likely to have problems in school with academics, classroom behavior, or both; when possible, clinicians undertaking a diagnostic evaluation for ADHD should conduct classroom observations in addition to interviews with teachers (Barkley, 2015). Although observational data may have limited diagnostic utility, they are often more useful than information collected from parent rating scales to inform goals for treatment if indicated (Pelham et al., 2005). Clinicians may also wish to screen for problems with adaptive, cognitive, and academic functioning to rule out intellectual disability or specific learning disabilities (Barkley, 2015). If pharmacological treatments are anticipated or family history is significant for ADHD or

other neurodevelopmental disorders, an updated physical exam, neuroimaging, or genetic testing may be warranted (AACAP, 2007; Barkley, 2015). After a diagnosis is made, periodic assessment using many of the same methods described above is recommended to determine change after interventions are implemented (Barkley, 2015).

Several challenges may arise when considering a diagnosis of ADHD. There is a striking difference in ADHD diagnoses by gender; boys outnumber girls by about two to one (APA, 2013). When girls are diagnosed with ADHD, they more often receive diagnoses of ADHD with predominantly inattentive presentation (APA, 2013; Rucklidge, 2010). The possible implication of this is that many girls go undiagnosed because they do not exhibit disruptive behavior at the same rate or intensity as their male counterparts who receive diagnoses of combined presentation or predominantly hyperactive/impulsive presentation (Rucklidge, 2010). Other demographic factors appear to increase the likelihood of receiving an ADHD diagnosis in the U.S. (Danielson et al., 2018). These factors include non-Hispanic ethnicity, speaking English as a primary language, living in a low-income household, and living in rural areas in the Midwest and South (Danielson et al., 2018). Notably, Danielson and colleagues (2018) found that Black children were more likely to receive an ADHD diagnosis than their White peers, a finding that departs from previous studies that have demonstrated under-identification and under-treatment of ADHD in Black children (e.g., Coker et al., 2016). These discrepancies highlight continued challenges in providing equitable and sensitive diagnostic and intervention services to children with ADHD in the U.S.

ASD Symptoms and Diagnostic Process

ASD is characterized by symptoms in two areas: social deficits and restricted, repetitive behavior (APA, 2013). Social criteria describe persistent challenges with social communication

and interaction across contexts (APA, 2013). These problems might take the form of limited social reciprocity, decreased awareness of nonverbal communication, and challenges developing and maintaining relationships (APA, 2013). In the domain of behavior, children with ASD demonstrate repetitive motor movements or speech, rigidity, restricted interests that are abnormal in intensity or focus, or sensory problems including seeking sensory input from the environment (APA, 2013). Specifiers include severity in the form of a rating for both social behavior and restricted, repetitive behavior ranging from Level 1 corresponding with "requiring support" (p. 52) to Level 3 corresponding with "requiring very substantial support" (APA, 2013, p. 52). Clinicians also specify whether the diagnosis of ASD is accompanied by intellectual impairment, language impairment, a known medical condition or environmental factor, another mental or behavioral disorder, or catatonia (APA, 2013). Symptoms of ASD cannot be better explained by intellectual disability or global developmental delay (APA, 2013). ASD symptoms must also be present in the early developmental period, usually around or before 24 months of age (APA, 2013).

Identification of ASD often occurs in two stages: screenings completed by pediatricians followed by full diagnostic evaluations. At the screening stage, the American Academy of Pediatrics (AAP) recommends that primary care doctors ask parents about developmental progress at all visits and conduct specific screenings for ASD symptoms at 18 and 24 months of age (Hyman et al., 2020). Hyman et al. (2020) noted that the typical process of developmental screening that occurs at 9-, 18-, and 30-month visits does not usually include questions about social development, emphasizing the importance of using ASD-specific screening tools in toddlerhood. The most common and robust parent rating scales used for screening are the Modified Checklist for Autism in Toddlers, Revised with Follow-up (M-CHAT-R/F; Robins et

al., 2009) and Social Communication Questionnaire (SCQ; Rutter, Bailey, et al., 2003). The Screening Tool for Autism in Toddlers and Young Children (STAT; Stone et al., 2000) is a brief, clinician-administered assessment that can also help to identify concerns about ASD. These screening measures can identify which children may be at risk of developing ASD and thus require further evaluation.

Universal screening for ASD in pediatric practices is not recommended past the age of 30 months, and no screening tools currently exist for children older than 30 months (Hyman et al., 2020). However, the SCQ (Rutter, Bailey, et al., 2003), the SRS-2, and the Children's Communication Checklist, Second Edition (Bishop, 2006) have evidence of strong to moderate construct validity among school-age children who had or later received diagnoses of ASD (Charman et al., 2007). Charman and Gotham (2013) commented that adequate screening for ASD, while important, certainly has limitations related to characteristics of samples used to create and establish psychometric properties of various measures. Clinicians who conduct universal screening for ASD ought to familiarize themselves with the sensitivity and specificity of available rating scales (Charman & Gotham, 2013).

Children who are identified as at risk for ASD should be referred to clinical, early intervention, or school services for further evaluation (Hyman et al., 2020). To allow families to access intervention services quickly, primary care clinicians may make an initial diagnosis of ASD using data from rating scales such as the BASC-3 (Reynolds & Kamphaus, 2015); the SRS-2; the SCQ (Rutter, Bailey, et al., 2003); the Diagnostic Interview for Social and Communication Disorder (Wing et al., 2002); or the CBCL. Whenever possible, these rating scales should be completed by multiple raters, most often by at least one parent and one teacher (Williams White et al., 2007). In more specialized settings, a focused caregiver interview such as the ADI-R or a

structured observation protocol, most often the ADOS-2 or the Childhood Autism Rating Scale, Second Edition (CARS-2; Schopler et al., 2010), may be used. These structured interview and observation tools are crucial datapoints in what are often referred to as "gold standard" multidisciplinary assessment practices for ASD (Falkmer et al., 2013). While these measures have evidence of strong technical adequacy, they should be used in combination with clinical judgement exercised by multidisciplinary teams (Charman & Gotham, 2013). In addition to evaluating core social and behavioral components of ASD, clinicians should strongly consider assessments in cognitive, language, adaptive, motor, and sensory domains (Hyman et al., 2020). Genetic testing, neuroimaging, and other medical tests may also be of interest to families and providers regarding possible risk factors and prognosis (Hyman et al., 2020).

In summary, diagnosis of ASD involves a combination of interviews, rating scales completed by multiple parties such as caregivers and teachers, and direct behavior observations completed by a trained clinician (Hyman et al., 2020). In ideal practice, this information is reviewed by a multidisciplinary team of clinicians, at which time a diagnostic decision about ASD can be made (Falkmer et al., 2013). Once a diagnosis has been determined and interventions are implemented, a child's progress is typically assessed using a single-case design in which data on target behaviors are collected frequently to determine the extent of change over time (Roane et al., 2011). The clinical benefit of monitoring progress with single-case design is that it allows children to be measured against themselves, which is especially useful considering the heterogeneity within the diagnostic category of ASD (Fisher et al., 2011).

Several factors can obscure appropriate identification of risk for developing ASD. The first is typical or advanced cognitive ability; many children who are later diagnosed with ASD exhibited advanced language, problem solving, or academic skills for their young age, thus

making it difficult to detect core symptoms of ASD until they reach school-age (Hyman et al., 2020). In addition, interference of similar disorders, specifically ADHD, can make it difficult to tease apart ASD symptoms (Miodovnik et al., 2015). While there is evidence that the diagnostic specificity of the SRS-2 is strong, a substantial portion of children with ADHD alone also meet criteria for ASD using the SRS-2 (Joshi et al., 2011; Reiersen et al., 2007). This may be due to shared problems with social skills or externalizing behavior (e.g., Cooper et al., 2014; Factor et al., 2017) and related challenges with creating measures robust enough to meaningfully differentiate symptoms of ASD and ADHD. These challenges are not specific to the use of the SRS-2. In one large study, 20% of parents who had children with an ASD diagnosis reported receiving a diagnosis of ADHD first, and the vast majority of these children did not receive an ASD diagnosis until approximately three years after their ADHD diagnosis, regardless of ASD symptom severity (Miodovnik et al., 2015). Externalizing behavior and low IQ scores may also reduce the specificity of screening and diagnostic tools (Charman et al., 2007).

Female gender is another complicating factor for accurate diagnosis, especially in the absence of intellectual impairment; among individuals with ASD, boys outnumber girls four to one (APA, 2013). Leading theories with varied empirical support include the possibility of a female-specific ASD phenotype with less severe symptoms (Mandy et al., 2011), a conceptualization of ASD that has historically focused on boys and men that has resulted in diagnostic criteria and assessment tools that underrepresent expressions of ASD in girls and women (van Wijngaarden-Cremers et al., 2014), underreporting or mislabeling symptoms due to behavioral expectations of girls or variations in behavior at school versus at home (Bargiela et al., 2016), and a phenomenon called “masking” or “camouflaging” in which girls intentionally hide symptoms of ASD in an effort to appear “normal,” especially around peers (e.g., Baldwin &

Costley, 2015). Whatever the reason for gender differences in diagnosis of ASD, lack of support associated with inappropriate or nonexistent diagnoses can lead to intersectional, long-term vulnerabilities. For example, a startling majority of one sample of women who received diagnoses of ASD in adolescence or adulthood reported histories of sexual abuse, attributing these instances to reliance on mimicking social behaviors, difficulty understanding social situations, and peer rejection (Bargiela et al., 2016). There are also disparities in the timing of early evaluation and ASD diagnosis of Black and White children, despite nearly equivalent prevalence, that interfere with access to early intervention services (Maenner et al., 2020). Black and Hispanic children who receive diagnoses of ASD are also more likely to be classified as having accompanying intellectual impairment compared to White counterparts (Maenner et al., 2020). These issues point to a need for increased access to early screening and diagnostic services, and a thorough examination of the tools used to identify ASD and its common correlates (Hyman et al., 2020).

Complications for Diagnosis of Comorbid ADHD and ASD

Severe Symptom Presentation. Children with comorbid ADHD and ASD often demonstrate impairment in more symptom domains than children diagnosed with either disorder alone. Children with comorbidity are likely to demonstrate poorer social skills (Dellapiazza et al., 2021; Factor et al., 2017; Rao & Landa, 2014) compared to children with ASD alone, and executive functioning and related externalizing behavior comparable to or worse than children with ADHD alone (Craig et al., 2016). Socially, children with comorbid ADHD and ASD demonstrated atypical facial processing on par with children with ASD alone (Groom et al., 2017), and made fewer gains during a social skills intervention than their peers with ASD alone (Antshel et al., 2011).

The impacts of comorbidity are also evident through externalizing behavior; children who only received a diagnosis of ADHD but exhibited symptom overlap with ASD tended to show more oppositional behavior and conduct problems (Cooper et al., 2014). A substantial body of work has corroborated the finding that children with comorbid ADHD and ASD tend to demonstrate increased externalizing behavior and lasting conduct problems compared to children with only one diagnosis (Craig et al., 2015; Flouri et al., 2015; Goldin et al., 2013; Jang et al., 2013). Other problems associated with comorbid ADHD and ASD may include lower full-scale IQ and working memory scores (Cooper et al., 2014; Rao & Landa, 2014), motor problems (Cooper et al., 2014), and lower adaptive functioning (Ashwood et al., 2015; Craig et al., 2015; Rao & Landa, 2014) compared to peers with only one diagnosis. Further, children with comorbid ADHD and ASD are at greater risk for developing internalizing and somatic symptoms (Cooper et al., 2014; Wilson et al., 2014).

In short, published literature overwhelmingly suggests that children who qualify for both ADHD and ASD diagnoses tend to display impacted forms of both disorders in more domains of development than children with only one diagnosis. However, it is often challenging in practice to differentiate ASD from ADHD; common assessment tools such as the ADHD Rating Scale-IV (DuPaul et al., 1998), ADI-R, and ADOS-2 have demonstrated limited adequacy in discriminating between ADHD and ASD (Grzadzinski et al., 2016; Yerys et al., 2016). Practitioners should explore assessment methods and tools carefully to determine both ADHD and ASD symptoms if either of these diagnoses is suspected (Grzadzinski et al., 2011; May et al., 2018). Comorbid presentations of ADHD and ASD warrant further attention in the literature; it is clear that children who have symptoms of both ADHD and ASD have poorer trajectories across the lifespan, yet there are unanswered questions about how best to assess and support children

who meet criteria for both (Antshel & Russo, 2019; Davis & Kollins, 2012; Leitner, 2014; Reiersen & Todd, 2008). More research is needed to provide appropriate guidance to practitioners who consider these diagnoses.

Challenges with Diagnostic Conceptualization. In addition to considerations for diagnostic practices, the conceptualization of ADHD and ASD has led to debate among clinicians and researchers about how these disorders relate to one another. Some authors believe that ADHD and ASD exist on a continuum, with relatively pure ADHD on one end and severe ASD on the other (Kern et al., 2012; van der Meer et al., 2012), and they point to shared genetic etiologies and neurological features. For example, one large-scale twin study found that communication problems commonly seen in children with ASD had a strong genetic relationship to children with ADHD (Taylor et al., 2015). van der Meer and colleagues (2012) demonstrated that ADHD and ASD symptoms can be categorized into five overlapping classes, with two of these classes exhibiting symptoms of both disorders but manifesting in one more prominently than the other (i.e., "ADHD[+ASD]" and "ASD[+ADHD]" [p. 1164]).

Related, some authors have published findings supporting the idea that ADHD and ASD exist on separate spectrums that overlap in the domain of executive dysfunction; Craig et al. (2016) concluded that children with ASD demonstrated more impaired flexibility and planning, children with ADHD demonstrated worse response inhibition, and children with both disorders demonstrated deficits in all of these areas of executive functioning. Therefore, comorbidity may not be a unique condition or phenotype, but an additive effect of symptoms of both ADHD and ASD (Craig et al., 2015, 2016). Other authors do not believe that ADHD and ASD exist on a spectrum, but rather that ASD naturally subsumes the core symptoms of ADHD and they conclude that ADHD should be considered a core feature of ASD (Mayes et al., 2012). However,

evidence also exists to suggest that large proportions of samples of children with only an ADHD diagnosis met criteria for ASD using the ADOS-2, suggesting that ADHD may also subsume characteristics of ASD in a substantial number of cases (Grzadzinski et al., 2016).

Conversely, a body of work views ADHD and ASD as clearly distinct (Green et al., 2016; Rommelse et al., 2018). Some authors have suggested that the suitability of applying both or just one diagnosis can be distinguished by examining the degree to which symptoms of both disorders cause impairments in daily life and thus require specific interventions (Rommelse et al., 2018). Rommelse and colleagues (2018) outlined a number of "red flags" (p. 679) in the early developmental period to support practitioners in identifying ASD over ADHD, including limited social development, sensory over- or under-reactivity, restricted interests, and low physical activity. In general, Rommelse et al. (2018) recommended that practitioners alternate between an "ADHD perspective" and "ASD perspective" (p. 679) to determine which represents the best fit for the needs of individual children. Although the nature of the relationship between ADHD and ASD is still contested, the possibility of shared etiology, overlapping domains of impairment, and implications for intervention services are areas that require further inquiry that are rooted initially in assessment practices.

Measurement of Symptoms and Impairment

Parent rating scales and direct behavioral observations are likely to play a role in the identification and monitoring of social skills and externalizing behavior of school-age children with ADHD and children with ASD (Barkley, 2015; Hyman et al., 2020). Using data from these sources in combination allows clinicians and researchers to gather information about behavioral manifestations of ADHD and ASD across environments and raters, ultimately providing a richer picture of development over time (Barkley, 2015; Hyman et al., 2020). Careful measurement of

social skills and externalizing behavior is important for both clinical and research purposes; selection of measurement methods can influence diagnostic decisions (Valo & Tannock, 2010) and conclusions drawn by researchers (Radley et al., 2015). To responsibly interpret and present research findings, a nuanced and thorough understanding of the strengths and limitations of these data collection tools is paramount.

A discussion about measurement tools and techniques is also relevant from a clinical perspective. Even during the diagnostic process which often results in access or lack thereof to intervention groups, common assessment tools have demonstrated poor discriminant validity between ADHD and ASD (Grzadzinski et al., 2016; Yerys et al., 2016), raising questions about what may or may not be most appropriate to evaluate outcomes when children with differing diagnoses or comorbid ADHD and ASD are combined for interventions. Before discussing which interventions are considered evidence-based for children with ADHD and children with ASD, it is helpful to examine methods and properties of commonly used measurement tools.

Utility of Parent Rating Scales

Rating scales are they are nearly ubiquitous in clinical settings as common assessment tools used in clinical and research contexts to identify and quantify domains and degree of impairment associated with ADHD and ASD (Barkley, 2015; Hyman et al., 2020; Ramirez & Sajatovic, 2012). Rating scales are popular instruments due to their efficiency in gathering information about symptoms and impairments from multiple informants who know the child well (Barkley, 2015; Tatsuoka, 2012). Rating scales uniquely contribute information about the statistical deviance of a behavioral or social-emotional presenting problem (Barkley, 2015). Moreover, most behaviors differ in frequency, intensity, and duration across contexts, and this variability is expected by clinicians (Konold et al., 2019). By using rating scales to collect

information about the topography of behavior from informants who observe the child in different contexts with different demands placed on them, it becomes possible to construct a more complete and meaningful picture of the child. The SRS-2 and CBCL are two such measures that can be completed by both teachers and parents to assess specific domains of social and externalizing behavior for children with ADHD and children with ASD. However, rating scales used without other methods of corroborating the nature of social and externalizing behaviors are vulnerable to a number of limitations, including measurement errors that can arise for a variety of reasons ranging from the size of and representation of the norm sample (Boateng et al., 2018) to biases held by the respondent (e.g., Stokes et al., 2011).

Utility of Direct Behavior Observations

Data collected from direct behavioral observations often inform the need for and response to intervention, especially for children with ASD (Roane et al., 2011). Applied behavior analysis (ABA) interventions often use single-case designs to monitor progress in which behavioral data are frequently collected during intervention sessions (Fisher et al., 2011; Roane et al., 2011). While direct observations are used less often when treating children with ADHD compared to children with ASD (Barkley, 2015), they may still be used to gather information about specific behaviors to increase confidence in treatment outcomes (Hoza et al., 2007). Continuous methods of recording such as frequency counts are the most robust and the most resource intensive (Thompson & Borrero, 2011). Data collection techniques such as partial interval recording (PIR) that do not require constant observation are often less accurate than continuous methods of recording, but the use of an interval system may be more efficient and better describe the topography of certain behaviors of interest (Thompson & Borrero, 2011). In addition, there is a strong precedent in research of using interval systems to record prosocial, externalizing, and on-

task behaviors (Lane & Ledford, 2014). The goodness of fit between the target behavior, the resources available to collect the data, and the selected observation system is of critical importance in research and clinical work (Thompson & Borrero, 2011).

Partial Interval Recording. Partial interval recording (PIR) is the most common method of data collection in ABA in which a defined behavior is recorded if it occurs during a pre-determined length of time (Bailey & Burch, 2018; Lane & Ledford, 2014). PIR is well-suited to behaviors with unclear beginnings and endings, and less intensive than and about as accurate as continuous recording of behavior (Sam et al., 2015). Smaller intervals and longer observation periods may result in reduced error, with intervals of three minutes or less demonstrating evidence of the strongest correspondence with continuous recording methods (LeBlanc et al., 2019; Wirth et al., 2014). However, even observation periods as brief as 10 minutes can be sensitive to changes in behavior if the intervals are short; in other words, a 10-minute observation with 10-second intervals is about as sensitive to change as a 30-minute observation with 30-second intervals (Devine et al., 2011). PIR is considered the best estimate of frequency among common interval systems but it is not recommended as a replacement for recording the frequency of behaviors (Lane & Ledford, 2014). PIR can underestimate high-rate behaviors (Repp et al., 1976; SungWoo et al., 2011), and overestimate durations of behavior (Gardenier et al., 2004; SungWoo et al., 2011; Wirth et al., 2014), in some cases resulting in overestimation of effect sizes of interventions (Radley et al., 2015). Errors related to inaccuracies in measurement of duration can be further minimized by designing the length of intervals to approximate the duration of the target behavior (Wirth et al., 2014).

Frequency Recording. Frequency recording refers to a continuous observation system in which a behavior is recorded every time it occurs during an observation period (Bailey & Burch,

2018; SungWoo et al., 2011). Frequency recording is best used for behaviors with clear beginnings and endings with consistent durations (Bailey & Burch, 2018; SungWoo et al., 2011). Frequency recording can be a reliable observation system for up to six different behaviors simultaneously (Bailey & Burch, 2018). To facilitate accurate comparisons across observations, observations must be of equal length; when this is not possible, it is recommended that data be converted to a response rate that specifies the frequency of the target behavior per a consistent time interval (SungWoo et al., 2011). Occasionally, frequency recording may be combined with short intervals to obtain the rate of a behavior; this system can result in comparatively low interobserver agreement (Bailey & Burch, 2018). In general, continuous recording methods demonstrate low interobserver agreement when documenting behaviors with long durations (e.g., participation for long stretches of an activity, tantrum behavior) compared to short or instantaneous behaviors (Rapp et al., 2011).

Using Rating Scales and Observations Together

Rating scales and direct behavior observations have respective benefits and limitations. For example, Williams White et al. (2006) noted that using parent rating scales to assess social skills of children with ASD allows researchers to address skill generalizability in a way that would be inefficient or impossible with behavioral observations alone. However, Williams White et al. (2006) also commented that behavior observation methods can be more appealing than rating scales due to reduced risk of observer/respondent bias, but challenges unique to direct behavioral observation methods include expense, effort, and technical problems with recording equipment. While rating scales are often less resource-intensive than direct observations (Kelley et al., 2011), they are likely not as sensitive to change during intervention because many scales may be too broad to capture changes in specific behaviors (Volpe & Gadow, 2010). For purposes

of progress monitoring, the trade-off of acceptability and rigor should influence the selection of measurement tools as the acceptability of data collection methods is often related to the quality of the data (Volpe & Gadow, 2010). Requiring caregivers to complete unabridged rating scales with many items on multiple occasions over the course of an intervention can burden respondents and diminish the quality of data collected (Volpe & Gadow, 2010).

For children with ADHD and children with ASD, unique challenges influence the clarity of data collected through parent rating scales. Considering the strong hereditary component of ADHD, levels of parental symptomatology can influence rating scale responses (Sayal & Taylor, 2005). Parents may also over-report symptoms of ADHD compared to other informants (Re & Cornoldi, 2009), or demonstrate limited sensitivity to ADHD symptoms compared to teachers on teacher rating scales (Tripp et al., 2006). In addition, child characteristics such as language ability, a common impairment among children with ASD (APA, 2013), can impact the degree of agreement between raters (Gooch et al., 2017). Taken together, these findings impress the importance of carefully selecting data collection methods and informants and choosing a combination of methods whenever possible. Rating scales play a unique role in conceptualizing problems, gathering data from multiple sources and settings, and periodic evaluation in intervention contexts; however, the utility of rating scales can be strengthened if used as one component of a multimethod assessment package (Barkley, 2015; Hyman et al., 2020).

From an applied behavior analytic perspective, direct behavioral observations are considered the “gold standard” (SungWoo et al., 2011, p. 119). Direct observations offer the added advantage of assessing aspects of the environment that influence the likelihood of target behaviors, and thus establish a functional relationship between stimuli and behaviors (Kelley et al., 2011). Among the most important considerations for the use of behavioral data in clinical

settings is the degree of resources necessary to obtain reliable data (e.g., Williams White et al., 2006). Behavioral observation methods require considerable time and effort to establish a reliable estimate of behaviors and are best-suited to observable behaviors, such as externalizing behaviors, that occur at moderate frequencies, as opposed to internalizing problems or behaviors that occur at low frequencies which may be better assessed using rating scales (Kelley et al., 2011; Volpe et al., 2009). In other words, direct behavioral observations may be less accurate than other measurement methods for specific target behaviors (Kelley et al., 2011; Volpe et al., 2009). Therefore, as with the use of parent rating scales, behavior observation methods are important measurement tools with limitations that can be addressed when combined with other methods of data collection (Bearss et al., 2013).

It is beneficial to use both indirect and direct measures of behavior related to ADHD and ASD for clinical purposes (Barkley, 2015; Kelley et al., 2011; SungWoo et al., 2011; Williams White et al., 2006). Indirect measures like rating scales do not always correspond with the behavior of interest and should be paired with direct observations whenever possible (SungWoo et al., 2011). However, making sense of parent ratings and direct measurement of behavior together can be complex; Sims and Lonigan (2012) examined the relationship between parent and teacher ratings of ADHD symptoms with a continuous performance task (e.g., presenting a series of stimuli and providing directions to quickly respond to one stimulus while not responding to others). They noted that, while parent and teacher rating scales of ADHD showed no significant correlation with one another, their respective relationships with the direct measure of the continuous performance task suggested that parent and teacher ratings may be indices of different aspects of ADHD evident in the child's behavior (Sims & Lonigan, 2012). Comprehensive progress monitoring systems that incorporate information about specific target

behaviors, such as behavioral data, as well as broader development, such as parent rating scales, are necessary for both identifying problems and measuring progress once an intervention has been implemented (Volpe & Gadow, 2010). In summation, using direct behavior observations and rating scales in combination yields a more detailed and robust measurement of externalizing behaviors and social skills for children with ADHD and children with ASD than either method can alone, and multimethod assessment during the course of intervention services supports trustworthy conclusions regarding intervention outcomes.

There are several considerations specific to conducting research with parent rating scales and behavioral observation methods. Direct behavioral measures may be attractive to researchers because they may be less vulnerable to bias and measurement error than indirect measures such as rating scales (Williams White et al., 2006). When using direct behavioral observations, researchers are likely to collect data on interobserver agreement to provide evidence that collected data are consistent across observers and observation sessions; the use of PIR may inflate interobserver agreement compared to continuous recording methods (Rapp et al., 2011). In addition, challenges with adequately measuring the duration of behavior using PIR can result in overestimated effect sizes of outcomes (Radley et al., 2015). Frequency recording is also vulnerable to problems with interobserver agreement when it is used to obtain the rate of a behavior (Bailey & Burch, 2018) or behaviors with long durations (Rapp et al., 2011).

Other methods of direct measurement used to investigate changes associated with interventions for children with ADHD and children with ASD may take the form of outcome measures. For example, Lopata et al. (2010) used both direct outcome measures and parent-report during a randomized control trial of a summer social skills intervention for children with ASD. The authors found that both scores obtained from parent rating scales and child

performance on outcome measures assessing verbal language, nonverbal communication, and social skills knowledge significantly improved (Lopata et al., 2010). Other studies have used neuropsychological tests to measure aspects of executive function in children with ADHD before and after participating in an intervention program (Gerber et al., 2012; Sims & Lonigan, 2012). This prior work provides support for the practice of using indirect and direct methods of data collection to evaluate intervention outcomes as both are useful and sensitive to change over time in an intensive treatment context for children with ADHD and children with ASD.

Previous work has also offered guidance on the use of parent rating scales for research purposes. For many intervention programs that target core problems of ADHD and ASD, a common goal is that skills learned in the context of the intervention will generalize to other environments such as home or after-school activities (Fabiano et al., 2014; National Research Council [NRC], 2001). Although it is impossible to collect consistent behavioral data for most children with ADHD and children with ASD in multiple environments, rating scales present an opportunity to gather data regarding broad behavioral and social outcomes that may be associated with participation in an intervention program (Barkley, 2015; Hyman et al., 2020). Therefore, rating scales provide one method by which researchers can examine generalization of skills, an important clinical concern for children with ADHD (Abikoff, 2009) and children with ASD (NRC, 2001).

Analyses of data collected from rating scales may be influenced by which constructs are measured by the scale itself. In a review of studies examining children with ADHD, Fabiano et al. (2009) noted that parent report of impairment yielded larger effect sizes than parent report of ADHD symptoms. In other words, parent-report measures that ask about independently following routines, problems with peers, or the nature of impulsive behaviors are likely to be

more sensitive to change over time than measures that seek to establish the presence or absence of DSM-5 criteria of ADHD (Fabiano et al., 2009). This finding is unsurprising considering that symptoms of neurodevelopmental disorders tend to be relatively intransigent where degree of impairment is variable (APA, 2013), but it provides guidance regarding how to best frame the constructs that are chosen as variables in research in intervention settings. However, even when the constructs of interest are specific and related to impairment, the use of parent rating scales may still impact conclusions drawn by researchers (Cordier et al., 2018). In a meta-analysis of peer-inclusion interventions for children with ADHD, Cordier et al. (2018) found that using parent rating scales as the primary outcome measure of social skills significantly mediated the treatment effect. In addition to these challenges, there are limitations to the interpretation of scores obtained from rating scales. As mentioned earlier, rating scale scores of children with ASD and externalizing behavior problems can obscure identification of social skills problems on measures such as the SRS-2 that are explicitly designed to measure social skills among children with ASD (Hus et al., 2013). These limitations must be accounted for when conducting research using data obtained from parent rating scales.

Evidence-Based Intervention Practices to Support Externalizing Behavior and Social Skills

As mentioned previously, ADHD and ASD share many common features of symptom expression, particularly in the areas of social and behavioral development (e.g., Barkley, 2016; Demopoulos et al., 2013; NRC, 2001). The overlap of etiology and symptom expression, as well as challenges with measurement of behavior, raises questions about appropriate intervention and whether children with comorbid diagnoses should be directed to different treatment pathways than counterparts with only one diagnosis, or if a combination of interventions that are employed with both ADHD and ASD populations is appropriate. There is evidence to suggest that the

presence of comorbid ADHD and ASD is associated with worse treatment outcomes on interventions specific to one disorder (Antshel et al., 2011; Reierson & Todd, 2008), but there is little evidence to date to support or discourage particular intervention practices for comorbid groups, or even the applicability of specific interventions for use across both ADHD and ASD populations (e.g., Antshel et al., 2016; Davis & Kollins, 2012; Smith et al., 2016). More research is needed to better identify unique challenges associated with comorbid ADHD and ASD and the most efficient clinical pathways from diagnosis to effective intervention.

However, studies that investigate social and behavioral outcomes of psychosocial interventions and their underlying principles nearly always include participants divided by ADHD or ASD diagnosis (e.g., Davis & Kollins, 2012). This distinction has created two primary problems for the literature base of evidence-based practices that address social skills and externalizing behavior in school-age children. First, limited research exists that can reasonably demonstrate utility of interventions for both children with ADHD and children with ASD, even when the clinical presentation may warrant a combination of approaches (Leitner, 2014). This is likely to lead practitioners to separate children with ADHD and children with ASD for psychosocial intervention, which may be both unnecessary and inefficient. Second, the bulk of intervention research regarding social skills and externalizing behavior has focused on children with either ADHD or ASD, and little research exists to inform intervention practices for children with comorbid ADHD and ASD.

Children ages 6-12 with diagnoses of ADHD or ASD represent one well-defined age group with which to study psychosocial interventions. The mean age of diagnosis of ADHD is 6 years and 2 months (National Institute of Mental Health, 2017). Although ASD is often diagnosed much younger with a reported median age of diagnosis of 51 months (Maenner et al.,

2020), ages 6-12 represent the developmental period between the time when ADHD is often reliably diagnosed and the beginning of adolescence. Psychosocial interventions for adolescents are likely to use different strategies and incorporate more advanced content than those used for middle childhood, as adolescence is a unique period of development that may require different or new supports to address challenges related to ADHD and ASD (Evans, Langberg et al., 2014; Levy et al., 2011). Therefore, it may be inappropriate to include adolescents in the same sample with children as young as 6 years old. For this reason, the following review of evidence-based interventions is limited to children between the ages of 6 and 12.

Evidence-Based Interventions in the ADHD Literature

Research investigating intervention programs and strategies for social development of children with ADHD have identified several practices that increase social skills and decrease externalizing behaviors. A foundational study on intervention strategies for children with ADHD was the Multimodal Treatment of ADHD (MTA) study (MTA Cooperative Group, 1999). Undertaken over the course of 14 months, this multi-site study compared randomly assigned treatment conditions among 579 children with ADHD: medication alone, behavioral treatment alone, combined medication and behavioral intervention, and community care (MTA Cooperative Group, 1999). Behavioral conditions included parent trainings, school-based interventions such as teacher consultations and school-home notes (sometimes also referred to as daily report cards), and contingency management (Well et al., 2000).

During summer breaks from school, children in the MTA study participated in a manualized intervention called the Summer Treatment Program (STP; Pelham et al., 2012), an intensive camp program. A brief description of the STP offers insight into the behavioral intervention strategies used in the MTA study. First, in the STP staff were present in a two-to-one

child-to-adult ratio and participated in extensive training prior to camp (Pelham et al., 2012). Behavioral data were collected for every child across the camp day to monitor progress (Pelham et al., 2012). The intervention component administered in the highest dose was a contingency management system that continuously awarded and deducted points based on behavior (Pelham et al., 2012). Chronis et al. (2004) and Coles et al. (2005) demonstrated that behavior worsened when this point system was withdrawn. In addition to this behavior management system, social skills instruction was emphasized throughout daily recreational activities through discussion, modeling, role play, in vivo practice, and social reinforcement (Pelham et al., 2012).

The STP also incorporated home collaboration through daily report cards and parent training meant to support long-term generalization of gains made during the intensive intervention to other environments, be it home, school, or extracurricular activities (Chronis et al., 2006; Pelham & Fabiano, 2008; Pelham et al., 2012). Importantly, the STP took place in a naturalistic environment with typical camp activities, such as sports games, swimming, and arts and crafts (Pelham et al., 2012). The STP is a well-studied intervention in its own right; over two decades of findings using samples of children ages 5-12 with ADHD from the MTA study and other sites have suggested positive outcomes for peer relationships, on-task behavior, and challenging behaviors (Chronis et al., 2004; Pelham & Fabiano, 2008; Pelham et al., 2000), as well as increased knowledge of sports rules, athletic skills, and improved sportsmanship (O'Connor et al., 2013).

The main findings of the MTA study suggested that the combined treatment group that received both medication management and behavioral intervention demonstrated the greatest improvement in ADHD symptoms over time (MTA Cooperative Group, 1999). However, children in the behavior-only condition also demonstrated meaningful improvements in peer

relationships and oppositional behavior, indicating that, while medication management can be useful for many families, behavioral treatments provide a unique opportunity to teach skills that support the development and functioning of school-age children (MTA Cooperative Group, 1999). A moderator analysis of outcomes at the end of the MTA study by Owens et al. (2003) found that high rates of parental depression as well as high severity of ADHD symptoms were associated with reduced change at the end of treatment within the medication-only and combination groups. These moderators were not significant for the other treatment groups, behavior-only, and community care (Owens et al., 2003). To interpret the finding that parental depression is a meaningful variable in treatment outcome, Owens and colleagues (2003) conducted supplementary analyses for the combination and medication-only groups and found that parental depression remained significant for the medication-only treatment group. The authors presented the hypothesis that in the combination treatment condition parents were also being treated, thus increasing involvement and efficacy (Owens et al., 2003).

At a follow-up 10 months after the conclusion of the treatment phase of the MTA study, gains made by the medication-only and combination groups reduced by about half; children in behavioral-only and community care conditions did not show the same magnitude of deterioration (MTA Cooperative Group, 2004a). Children in the medication-only group tended to take higher doses of medication than those in the combined group and may have experienced more severe side effects than their counterparts in other conditions (MTA Cooperative Group, 2004a). Children who were consistent in their use of medication or who began using medication in the interim between treatment and follow up showed the least deterioration, but less physical growth compared to peers (MTA Cooperative Group, 2004b). In contrast, a sizeable portion of the behavior-only group was not on medication at follow up, despite recommendations to begin

medication at the conclusion of treatment (MTA Cooperative Group, 2004a). Because medication treatment is so prevalent among children with ADHD, these factors may be meaningful when considering the full complement of intervention options based on family preferences and need. In another follow-up study eight years after initial treatment, participants were screened for negative school-based outcomes: truancy, discipline, and physical fights (Reed et al., 2017). The authors of this study acknowledged that only some childhood predictors were "modifiable" (p. 305), and suggested that by providing parents with knowledge of behavioral principles, engaging fathers in parent training, and equipping families with disciplinary strategies other than physical punishment, children were theoretically less likely to go on to develop negative behavioral outcomes as adolescents (Reed et al., 2017).

The MTA study and its follow-up studies have shaped the way practitioners intervene on behalf of children with ADHD. Several general principles can be distilled from the findings described above that inform intervention planning for children with ADHD. First, parent involvement and investment in treatment is critical to achieve the best possible outcomes. Children benefit when their parents are supported in learning skills to manage their child's behavior. Second, while medication can result in rapid improvement in symptoms and functioning for many children with ADHD, research suggests that pharmacological treatment is not a substitute for social skills instruction, compliance training, parent training, or behavioral management. Further, long-term use of stimulant medications to curb impulsive behaviors can have detrimental impacts on a child's overall health. Third, behavioral and social skills problems associated with ADHD must be targeted explicitly through consistent contingency management and other behavioral supports. Not only are behavioral interventions important for learning and

practicing skills, but they were also shown to reduce or even eliminate the need for long-term and high-dose use of medication treatment (MTA Cooperative Group, 2004a).

Since the publication of the results of the MTA study, research has continued to support the same psychosocial intervention strategies for children with ADHD. Literature reviews by Chronis et al. (2006), Evans, Owens et al. (2014), Fabiano et al. (2009), Fabiano et al. (2014), Pelham and Fabiano (2008), and Schoenfelder and Sasser (2016) identified behavioral parent training, classroom management, and behavioral peer interventions with contingency management (such as the STP) as the practices with the greatest degree of empirical support. Interventions have shown the most positive outcomes when they are sustained, intensive, and administered in naturalistic settings (Fabiano et al., 2014). It is noteworthy that in some of these reviews, weekly social skills groups, traditional social skills training (SST), and office visits with a practitioner were examined as strategies to address impairment related to ADHD and were found to lack empirical support (Evans, Owens et al., 2014; Pelham & Fabiano, 2008; Schoenfelder & Sasser, 2016). Evans, Owens et al. (2014) noted that much of the published literature has focused on increasing access to evidence-based practices, such as adapting parent trainings for fathers and single mothers and offering trainings over the phone.

Evidence-Based Interventions in the ASD Literature

In 2001, the NRC published a report that synthesized research regarding educational and clinical practices for children with ASD. A key feature of clinical practice highlighted in this work is the use of blends of individual teaching or behavioral strategies to comprise a comprehensive intervention package for children with ASD (NRC, 2001). The overwhelming majority of these comprehensive models such as the Early Start Denver Model (Dawson & Rogers, 2010) lean heavily on collections of strategies informed by ABA, which makes use of

teaching strategies such as pivotal response training, positive behavior supports, and functional behavior analysis (NRC, 2001). Following from this theory, the NRC (2001) identified 10 shared characteristics across empirically-supported intervention programs for children with ASD: (a) intervention starts early, (b) intervention is intensive in hours, (c) families are actively involved, (d) staff are highly trained, (e) progress is assessed on an ongoing basis, (f) instruction is systematic, (g) the environment is supportive, (h) communication and other impacted developmental areas are foci of interventions, (i) programs plan for generalization and maintenance, and (j) programs are appropriately individualized. Building upon the principles identified by the NRC (2001), Odom et al. (2010) identified 24 evidence-based practices for children with ASD including antecedent-based intervention; differential reinforcement of alternative, incompatible, or other behavior; functional behavior assessment; naturalistic intervention; prompting; reinforcement; social narratives; SST; and visual supports. In a later review, Wong et al. (2015) added several categories of evidence-based practices, including modeling and structured play groups, and offered continued support for the practices identified by Odom et al. (2010).

Most intervention practices for children with ASD are associated with ABA. According to Iadarola and Smith (2016), ABA is best described as an "overarching framework" (p. 256) that is based on observable behavior, individualized instruction, management of antecedents and consequences, and progress monitoring. ABA principles define and comprise entire treatment packages such as Early Intensive Behavioral Intervention (e.g., Lovass, 1981), Learning Experiences: An Alternative Program for Preschoolers and Parents (LEAP; e.g., Strain & Bovey, 2011), and the Early Start Denver Model (Dawson & Rogers, 2010), and can also refer to other collections of strategies assembled based on individual characteristics (Iadarola & Smith, 2016).

ABA strategies are considered the most effective behavioral intervention for individuals with ASD (Iadarola & Smith, 2016).

Evidence Base to Support Externalizing Behavior for Children with ASD. It is well-established that children with ASD often exhibit more externalizing behaviors than peers without identified disabilities (Bauminger et al., 2010). Importantly, functional behavior analysis is used to determine the function of challenging behaviors, or the reason a behavior occurs, which allows practitioners to design an intervention program that satisfies the function of the behavior in a more adaptive manner (Odom et al., 2010; Wong et al., 2015). Strategies used to reduce challenging behaviors often include antecedent and consequence manipulation, differential reinforcement, and other elements consistent with ABA (NRC, 2001). Externalizing behaviors among children with ASD are also often associated with communication impairments (e.g., Boonen et al., 2014) and sleep disturbance (e.g., Lindor et al., 2019); therefore, functional communication skills and resolution of sleep problems may be pivotal intervention targets. While few parent training programs for children with ASD focus on externalizing behavior (Davis & Kollins, 2012), those that exist tend to teach ABA and antecedent strategies to caregivers (Bearss et al., 2013; Sellinger & Elder, 2015) and emotion coaching (Wilson et al., 2013). Evidence from these studies indicates use of key strategies results in promising improvements in externalizing behavior among children with ASD.

Evidence Base to Support Social Skills for Children with ASD. For children with diagnoses of ASD, social skills impairment is a core diagnostic feature and therefore is necessarily a target of intervention (APA, 2013). Williams White and colleagues (2007) offered a review of available evidence to support social skills instruction practices. Strategies supported by Williams White et al.'s (2007) review included explicit definition of social rules, natural

reinforcers, reinforcement of response attempts, differential reinforcement, behavior tracking, peer and parent involvement, and practice opportunities in naturalistic environments such as peer-mediated interventions. Peer-mediated interventions specifically are associated with improved social contact in school and camp settings (Chang & Locke, 2019; Martinez et al., 2019; Odom, 2019). SST is also generally an effective practice, although methodological and theoretical differences across studies make the literature somewhat variable (Rao et al., 2008). Using the program Remaking Recess, Kretzman et al. (2015) conducted a small randomized control trial involving paraprofessionals providing encouragement and coaching on the playground during recess. Peer engagement increased for children with ASD during treatment (Kretzman et al., 2015). Lopata et al. (2007) used the Skillstreaming curriculum (Goldstein et al., 1997) in a summer camp setting to teach social skills for four hours per day to children with ASD without intellectual impairment; parent and staff ratings indicated improvement in social skills after treatment. Taken together, these findings indicate that empirically supported social skills interventions for children with ASD can take a variety of forms, but most involve naturalistic peer interactions and direct instruction from adults.

Differences and Similarities in Published Intervention Research for ADHD and ASD

Social Skills Training. The scientific body of evidence supporting psychosocial interventions targeting social skills and externalizing behavior for children with ADHD and children with ASD is distinguished by several meaningful differences. One practice, SST, appears to have the most distinct research support across diagnostic groups. SST is an established intervention practice for children with ASD (Odom et al., 2010; Wong et al., 2015), but has inconsistent empirical support among children with ADHD (Evans, Owens et al., 2014; Pelham & Fabiano, 2008). This is puzzling because children with ADHD have problems with

peers (Bagwell et al., 2001), and these problems may improve after participating in programs that incorporate SST (Chronis et al., 2004; Pelham et al., 2000; Pelham & Fabiano, 2008).

It is possible that the differences in effectiveness of SST are related to the core symptoms of the respective disorders; peer problems for children with ADHD may stem from impulsive behavior, while children with ASD may struggle to understand and replicate social cues and responses (APA, 2013). It may also be the case that social skills are not often described, studied, or measured appropriately among children with ADHD; much of the literature that has demonstrated limited efficacy of SST among children with ADHD has taken place in office settings, removed from opportunities to practice with peers (Mikami et al., 2017). One theory offered by Mikami et al. (2017) is that SST can be effective for children with ADHD if accompanied by strong social reinforcement, reminders of expectations, and coaching peers to be accepting of others. This finding suggests that perhaps SST is a viable intervention strategy for children with ADHD if executed in combination with other practices that support the success of children with ADHD. This conclusion is consistent with the findings of Chronis et al. (2004), Pelham et al., (2000), and Pelham and Fabiano (2008), who documented improved social skills in children with ADHD after participating in the STP, a comprehensive program that incorporates the principles Mikami et al. (2017) outlined. Interestingly, adult coaching of supportive peers as described by Mikami et al. (2017) is also a necessary ingredient in peer-mediated intervention models often used to increase peer social interaction among children with ASD (NRC, 2001).

Parent Support. Another noteworthy difference in intervention research among children with ADHD and children with ASD pointed out by Davis and Kollins (2012) is that, while approaches across diagnostic groups emphasize the importance of supporting parents, the role of the family is emphasized differently. For families parenting a child with ADHD, parent

components are often packaged as “parent training,” in which parents are taught how to manage the challenging behaviors of their children (Davis & Kollins, 2012). In contrast, families who parent children with ASD are often provided “parent education” that is more individualized and incorporates more information focused on supporting their child’s social skills (Davis & Kollins, 2012). This divergence in the conceptualization and content of parent involvement largely disregards any overlap in symptom presentations of children with ADHD and children with ASD. For example, although social skills problems are not a key feature of ADHD, it is clear that many children with ADHD struggle to make and keep friends (e.g., Bagwell et al., 2001) and therefore may benefit when parent interventions focus on supporting social skills and peer relationships. Similarly, many children with ASD demonstrate disruptive behaviors that are best targeted using a behavioral management approach commonly applied to children with ADHD (e.g., Bearss et al., 2013).

Research Methods. Finally, different dominant methodologies seem to define research into social skills and externalizing behavior interventions for children with ADHD and children with ASD. ASD by nature is heterogenous, making it a complex disorder to study and treat in group contexts. As a result, published research has predominantly used single-case designs to best accommodate the unique presentation of individual children (McGrew et al., 2016; NRC, 2001). McGrew et al. (2016) noted that while single-case designs may be clinically appropriate given the heterogeneity of ASD, over-representation of this design has likely limited the external validity of existing research. The lack of group-design studies also provides limited information about generalizability of interactions between individual characteristics and treatment effects (McGrew et al., 2016). Replicating interventions with multiple children is primarily how external validity is established within single-case research (NRC, 2001), making meta-analyses and

literature reviews a powerful way to consolidate information to make decisions about the relative efficacy of interventions targeting social skills and externalizing behavior for children with ASD. Kasari and Smith (2016) elaborated on the work of McGrew and colleagues (2016), emphasizing the need to identify "active ingredients" (p. 262) of targeted ASD interventions through group designs. Kasari and Smith (2016) advocated for a person-centered approach that acknowledges the complexities of individual presentations of ASD, comorbidities, culture and treatment preferences, and other characteristics until the field develops a better understanding of how to match children with distinct profiles to appropriate interventions.

By contrast, studies investigating social skills and externalizing behavior among children with ADHD more often take the form of randomized control trials, such as the MTA study, or some variation of a group design. It is possible that group designs are more clinically appropriate for the relatively more uniform nature of ADHD, thus bolstering the external validity of studies examining social skills and externalizing behavior interventions for children with ADHD. Just as there are a small number of randomized-control trials among ASD intervention literature (e.g., Lopata et al., 2010), single-case research also exists within the ADHD intervention literature, especially in studies in classroom settings (e.g., Harrison et al., 2019), but it does not appear to account for the majority of studies. The difference in methodology between ASD and ADHD research signals a qualitative difference in the kinds of empirical evidence, and associated methodological benefits and drawbacks, produced by existing research.

Commonalities in Intervention Literatures. Despite varied evidence for SST, parent programs, and research methodology, many practices with empirical support to improve social skills and externalizing behaviors for both children with ADHD and children with ASD are similar in clinical practice, if not indistinguishable. The most prevalent shared practice is an

emphasis on behaviorism and ABA. The STP (Pelham et al., 2012), for example, is at its core a collection of evidence-based practices based on ABA (Fabiano et al., 2014), the body of behavioral strategies with the largest evidence base of ameliorating symptoms of ASD (e.g., Iadarola & Smith, 2016; Wong et al., 2015). It is perhaps unsurprising, then, that the STP (Pelham et al., 2012) bears many commonalities with intervention features recommended by the NRC (2001), Odom et al. (2010), Williams White et al. (2007), and Wong et al. (2015), including high treatment intensity, treatment targets that are both comprehensive and individualized, clear structure, high child-to-staff ratio, consistent data collection, naturalistic setting, consideration of maintenance and generalization, prompting, social narratives, parent involvement, and constant reinforcement. It stands to reason that if ABA and the STP share a theoretical background as well as many evidence-based practices for children with ASD, programs like the STP may be associated with gains for children with ASD that are comparable to those that have been demonstrated for children with ADHD.

There is substantial overlap in the theoretical models that underpin most of what is considered evidence-based practice for social skills and externalizing behavior for children with ADHD and children with ASD. However, these theories have yet to be translated into clinical practice and research for comprehensive treatment approaches based on impairment rather than an ADHD or ASD diagnosis. This gap in the literature is most apparent when examining evidence to support intervention practices for children with comorbid diagnoses. While the volume of research documenting features and prevalence of comorbid presentations of ADHD and ASD has increased in recent years, intervention research has not demonstrated a comparable increase (Antshel et al., 2016). There are a number of studies comparing responses to psychopharmaceutical intervention across diagnostic categories and comorbid presentations

(e.g., Santosh et al., 2006), but published research has rarely documented the effects of psychosocial interventions on combined or comorbid samples (Davis & Kollins, 2012; Reirsen & Todd, 2008).

Psychosocial Interventions Supported for Children with ADHD and Children with ASD

A small number of studies summarize the bulk of psychosocial intervention research that has included participants with diagnoses of both ADHD and ASD. Papadopoulos and colleagues (2019) conducted a study of school-age children with comorbid diagnoses of ADHD and ASD that demonstrated the effectiveness of a behavioral intervention on sleep problems. Antshel et al. (2011) used parent rating scales to demonstrate the relative difficulty of achieving gains in social skills for children with comorbid ADHD and ASD compared to children with ASD and comorbid ASD and anxiety over the course of a 10-week program. The lack of available information has contributed to a lack of appropriate treatment; for example, Joshi et al. (2017) found that few psychiatrically-referred children with comorbid diagnoses of ADHD and ASD had ever received treatment, pharmacological or otherwise, for their ADHD symptoms. However, some studies are beginning to examine the effect of psychosocial interventions that have previously been established only for one diagnostic group, which has promising implications for practitioners and children with comorbid ADHD and ASD.

The STP (Pelham et al., 2012) is one intervention that incorporates many evidence-based practices across both the ADHD and ASD literature, although only a small number of studies have examined implementation of the STP for school-aged children with ASD. Two studies used single-case designs to investigate the potential benefits of participating in the STP for boys with high-functioning autism (HFASD); both studies concluded that the frequency of negative behaviors decreased while prosocial behaviors increased over the course of the STP (Mrug &

Hodgens, 2008; Wymbs et al., 2005). Mitchell and colleagues (2015) conducted the first quantitative group study to examine behavioral outcomes of 20 boys with HFASD who attended the STP. The authors documented significant improvements on measures of attention, on-topic contributions to group discussions, and following activity rules; they also noted a decrease in complaining and verbal abuse. While Mitchell et al. (2015) collected behavioral rating scales from parents and teachers prior to participation in the STP, these measures were not repeated after treatment or used in conjunction with behavioral data for analysis. One study by Dunne (2020) examined the factor structure and the possible growth models of behavioral data collected during a modified version of the STP that included both children with ADHD and children with ASD, referred to as Apex Summer Camp. Primary findings from this study supported a three-factor structure of negative behavioral data and a unidimensional structure of prosocial behavior, and suggested diagnosis and returner status (i.e., participating in the intervention for multiple summers in a row) as variables that predict change in behavior over time. One recent review of the literature by Sidol and Epstein (2020) identified gender; ADHD presentation; and comorbidities, especially callous-unemotional traits, as treatment moderators among children with ADHD who participated in the STP. Sidol and Epstein (2020) also highlighted a lack of published research regarding participation of children with ASD in the STP, attributing their exclusion to concerns that children with ASD would not benefit from strategies used in the STP; however, more evidence is needed to determine the validity and nature of these concerns.

A new development in this line of research is the use of the STP model among samples of preschool children (STP-PreK) with diagnoses of ASD, ADHD, or both. Initial findings suggest that, among children with ASD and externalizing behavior challenges, parents reported improvements in hyperactivity, inattention, aggression, social and adaptive skills, executive

functioning, and emotional regulation (Ros-Demarize & Graziano, 2021). Further, in a combined sample of children with ASD and children with ADHD who participated in the STP-PreK, Ros and Graziano (2020) found that ASD symptoms predicted membership in a high executive functioning deficit group while ADHD symptoms predicted membership in a high emotional regulation deficit group. Authors determined that membership in these groups predicted treatment response, and that the high executive functioning deficit group made the largest gains (i.e., the group whose membership was predicted by ASD symptoms). These findings are foundational not only to the idea that it may be possible to create symptom profiles prior to intervention that predict response, but also that relying on symptoms, not diagnosis, is a better way to approach intervention among children with ADHD and children with ASD (Ros & Graziano, 2020). Overall, these early studies of the use of the STP model among preschool children further support including children with ASD in this treatment context.

In summary, the body of research that contributes to an understanding of the treatment effects of the STP on school-aged children with ASD is promising, but much of it is limited to a small number of boys with HFASD. Moreover, none of the published prior works systematically evaluated parent rating scales to assess the success of children outside of the STP, severely limiting the field's understanding of how gains made during the STP can be meaningful in other areas of children's lives. More research is needed to expand the field's understanding of response to evidence-based interventions in clinical settings that incorporate children with ADHD and children with ASD, as well as how best to measure and study these changes.

Purpose and Research Questions

Despite foundational similarities in common evidence-based practices, there is a dearth of information in existing research literature regarding overlap of intervention practices for children

with ADHD and children with ASD. Further, because the relationship between these disorders is ill-defined beginning with the diagnostic process, more information is needed regarding measurement of social and behavioral impairments associated with ADHD and ASD, and outcomes associated with interventions that combine best practices of ADHD and ASD intervention literatures. The present study seeks to contribute to the understanding of psychosocial interventions with combined samples of children with ADHD and ASD by investigating social and behavioral outcomes of children with ADHD, children with ASD, and children with comorbid ADHD and ASD who attended the Apex STP. Moreover, this study also aims to provide a better understanding of the value of parent rating scales as predictors of behavioral outcomes for these children. Last but not least, because Apex is a modified version of the STP, which is a well-supported intervention for children with ADHD, the present study provides evidence regarding social and behavioral trajectories associated with participation in this modified treatment package. The specific research questions were as follows.

1. What is the average daily growth rate in children's prosocial and externalizing behaviors during an intensive summer intervention for children with ASD, ADHD, and comorbid ASD and ADHD?
2. To what extent does diagnosis predict growth in prosocial and externalizing behavior among children with ASD, ADHD, and comorbid ASD and ADHD who participated in an intensive summer intervention?
3. To what extent does parent-reported social and externalizing behavior predict changes in observed prosocial and externalizing behavior among children with ASD, ADHD, and comorbid ASD and ADHD who participated in an intensive summer intervention?

Chapter 3: Method

Participants

All participants in the present study are those who were accepted to participate in the 2019 Apex Summer Treatment Program (STP), which requires families to complete an online application and an in-person screening interview. The original sample comprised $N = 136$ children. Of this group, 18 participants were excluded: 15 were typically developing siblings of other enrolled children with diagnoses, had a diagnosis other than ASD or ADHD, or had only provisional diagnoses; one was excluded because their parent declined to consent to participate in research; and two were excluded because they were absent for the entire final week of camp. The final sample used in analyses thus included $N = 117$ children, who ranged in age from 5 to 12 years ($M = 8.12$ years, $SD = 1.47$). The sample was primarily White ($n = 77$; 65.8%), and mostly male ($n = 95$, 81%). Thirty-three percent ($n = 39$) had a diagnosis of ASD, 43% ($n = 50$) had a diagnosis of ADHD, and 24% ($n = 28$) had comorbid diagnoses that included both ASD and ADHD. Approximately 43% ($n = 50$) attended Apex in at least one previous year. Complete descriptive data are provided in Table 1.

Ten campers were missing parent rating scales. Of these, 40% were returners, and 90% were male. They had a mean age of 7.90 years ($SD = 1.45$). Sixty percent were White, 30% were Asian, and 10% were multiracial. Forty percent had a diagnosis of ASD, 30% had a diagnosis of ADHD, and 30% had comorbid diagnoses that included both ASD and ADHD. Asian campers were somewhat overrepresented in this group of campers with missing rating scales; otherwise, the demographics of this group were comparable to the larger sample.

Setting and Procedures

Data for the present study were collected during a naturalistic summer camp intervention for children with ASD, children with ADHD, children with comorbid ASD and ADHD, and typically developing siblings. Camp days were six hours per day, five days per week, and took place over the course of five weeks. Children were divided into small groups, referred to as bunks, based approximately on age or developmental level. In addition, considerations were made for placing female campers with other female campers in their age group. For example, in Bunk 1 there were 14 children aged 5 to 7 years that were all male, in Bunk 2 there were 14 children aged 6 to 7 years with six female campers and eight male campers, and so on.

Intervention Procedures

All children in this study participated in the Apex STP, an intervention modeled on STP (Pelham et al., 2012). The STP was developed originally as a naturalistic setting in which to provide intensive behavioral intervention for children with ADHD across eight weeks for nine hours per day. Apex took place over five weeks for six hours per day and incorporated many typical summer camp activities with several unique features. First, like the STP, campers earned points in a cost-response token economy system for exhibiting prosocial and compliant behavior, and campers lost points for exhibiting aggressive, noncompliant, and anti-social behaviors; this system allowed counselors to provide immediate feedback about behavior (Pelham et al., 2012). Points were exchanged for prizes at the end of each day (Pelham et al., 2012). Second, staff wrote individualized goals for each camper using daily report cards, and their average percentage on these earned privileges for Friday field trips to locations around the community; these privileges were independent of weekly point totals (Pelham et al., 2012). For example, a child who achieved an average of greater than 75% of their weekly goals would be on the highest field

trip level; an associated privilege might be extra tokens on a field trip to an arcade. A child who achieved an average between 50 and 75% of their weekly goals would earn some field trip privileges, while a child who achieved an average below 50% would have no additional privileges. In Apex, all campers attended field trips; this is unlike the STP where children who achieved less than 50% of their weekly goals remained at the camp site during field trips. Third, both the STP and Apex included instruction in social skills and opportunities to model and practice them throughout daily activities (Pelham et al., 2012). In addition to these therapeutic features for children, both the STP and Apex provided optional parent training sessions on a weekly basis.

Activities that took place during Apex included a skills drill in which children focused on specific aspects of playing the week's sport, a sports game, a counselors' choice activity, and board games, which replaced the academic learning center activity in the original STP. Unlike the original STP, swimming was not included in Apex. Each activity was highly structured and included pre- and post-activity discussions taken directly from the STP manual in which counselors reviewed rules, reminded campers of the week's social skill and solicited ideas about using the social skill in the activity, and informed children of points earned during the activity (Pelham et al., 2012). To support children's participation in these activities, Apex used a modified time-out procedure. The STP's time-out procedure can result in assigning a time out for up to 60 minutes in response to aggression, destruction of property, or repeated noncompliance. Apex used a modified time-out procedure in which a time out was assigned for a maximum of 15 minutes, without the opportunity to reduce time. During the year that data for the present study were collected, an additional modification to the time out procedure was piloted in which children who did not serve time outs appropriately were given the choice to complete the

remainder of the time out appropriately at camp or have a consequence at home (e.g., losing 20 minutes of screen time). Families were notified of any home consequences incurred via phone call.

Training Procedures

All Apex counselors were undergraduate students, graduate students, or professionals in a related field such as education, applied behavior analysis (ABA), occupational therapy, or speech therapy. Counselors participated in a two-week training prior to camp which included instruction, modeling, practice, and planning of essential intervention components. Counselors periodically completed and reviewed written assessments to determine their mastery of intervention content. Camp directors provided supervision and corrective feedback throughout training activities. Reliability and fidelity data were not collected during training.

Fidelity Procedures

During the camp, directors and head counselors completed at least one weekly fidelity check per bunk using a standardized protocol from the STP intervention manual that documented both intervention structure and process. One director or head counselor completed the entirety of each fidelity check. Fidelity checks took place during skills drills or sports games, which were split into four equal time intervals and began and ended with pre- and post-activity discussions. The fidelity check form included a checklist of pre-activity and post-activity group discussion elements such as reviewing the rules of group discussion. While the fidelity check form from the original STP stipulates that counselors announce that good sportsmanship is in effect at the end of the pre-activity discussion, Apex counselors evaluated children's sportsmanship from the beginning of the pre-activity discussion without this announcement; one item on the fidelity checklist in which counselors announce the beginning of good sportsmanship was therefore

disregarded. After documenting the structure of the discussion, the director or head counselor rated the counselor leading the group discussion on their pacing, reflection of child contributions, consideration of developmental level, body language, and tone of voice on a Likert scale (1 = "superior," 4 = "adequate," 7 = "inadequate").

During the following three intervals of the activity, the director or head counselor rated each counselor using the same 1-7 Likert scale (1 = "superior," 4 = "adequate," 7 = "inadequate") based on their perceived effectiveness in three domains: use of praise, coaching, and standardized attention questions (e.g., "What is the score and who is winning?"; see Figure 1). When evaluating coaching and standardized attention questions during the third and fourth interval respectively, directors and head counselors also completed short checklists of intervention structure (e.g., "Sport leader takes notes of child performance on a note card"). The last item on the fidelity checklist was an overall Likert scale rating between 1 and 7 (1 = "superior," 4 = "adequate," 7 = "inadequate") for each counselor's performance during the activity. Each fidelity check was followed by individualized feedback to counselors from the director or head counselor at the end of the same day.

Across all groups and staff members, 83% (n = 65) of group discussion elements, such as reviewing the rules of group discussion, were present during the sports activities observed. Counselor effectiveness of praise, coaching, and standardized attention questions were rated between "adequate" and "superior" in 93% of opportunities (n = 151) by directors and head counselors. To conclude that an evidence-based practice is implemented with fidelity, data should be collected on an ongoing basis that is representative of the intervention (i.e., 20% of intervention sessions; Breitenstein et al., 2010; Collier-Meek et al., 2013), and the data should reflect at least 80% fidelity (IRIS Center, 2014). Because each intervention group was observed

during 20% of the camp days and the evaluation of intervention elements and style was rated "adequate" or better in more than 80% of observations, there is sufficient evidence to conclude that the Apex was implemented with fidelity.

Interobserver Agreement Procedures

Interobserver agreement (IOA) was examined by comparing point counts of secondary observers with point counts of the primary point sheet counselor. IOA data were collected twice for every bunk. Point-by-point agreement was calculated for all observations. Across all observations, staff achieved an average of 87.80% agreement. It is recommended that IOA data are obtained frequently, usually in 20% of intervention activities (e.g., Horner et al., 2005); in this case, each group of counselors was only evaluated on two out of 24 days, or 8.33%, of the intervention program. Therefore, while the existing evidence to support strong IOA is promising, IOA data should be regarded with caution due to the low frequency at which they were collected.

Data Collection Procedures

Families provided demographic information and consent to participate in research as part of their application for Apex. In the two weeks prior to camp, parents completed several measures for each child they enrolled in the program, including the SRS-2 and CBCL. Item-level data were entered by research assistants into an Excel spreadsheet for scoring, then transferred to *SPSS* (Version 26) for cleaning and analysis of descriptive data. During the camp day, clinical staff monitored and recorded the behavior of each child attending camp according to operational definitions in the STP manual by using iPads to access cloud-based software specifically designed for Apex. These data were automatically populated into an Excel spreadsheet and imported into *SPSS* (Version 26) for cleaning and analysis of descriptive data. The final dataset was then imported into *R* (Version 4.0.2) for analysis.

Measures

Behavioral Data

Behavioral data were collected according to operational definitions in the STP manual. The full list of behavioral categories and their definitions is presented in Table 2. Behaviors were divided into interval and frequency categories. Intervals were defined by dividing the allotted time for any given activity into four equal segments; depending on the activity, intervals ranged in length from approximately seven minutes to 15 minutes. Following Activity Rules and Good Sportsmanship (when applicable) were scored according to intervals, meaning that if a child exhibited no instances of Violating Activity Rules or Poor Sportsmanship, respectively, for the duration of an interval, they were awarded points for these interval categories. After the first instance within an interval of Violating Activity Rules or Poor Sportsmanship, these interval categories effectively became frequency categories that then captured the number of instances of Violating Activity Rules and Poor Sportsmanship for the remainder of the interval. A third interval category, Behavior Bonus, was used to award children points for each interval in which they did not exhibit any negative frequency category behaviors. While these interval categories make up an important part of the STP intervention, they were excluded from this study as they were unlikely to directly capture the constructs of interest, prosocial behavior and externalizing behavior.

The remainder of the behavior categories were recorded as frequencies. For each instance of a behavior that met criteria for any frequency category, the child was informed of the relevant point gain or loss and the behavior was recorded. For continuous behaviors that met criteria for frequency category behaviors, a new instance was recorded for each minute of the behavior's duration. For example, if a child cried continuously for four minutes, the point sheet counselor

would record four instances of Complaining/Whining. Due to the large number of possible behavior categories of interest, previous researchers have used composites derived from these individual categories (e.g., Lopez-Williams et al., 2005; Mikami et al., 2010). For the purposes of the current study, several categories are of particular interest to comprise a social behavior composite and an externalizing behavior composite. To investigate changes in prosocial behavior, the following point system categories were used: Contributing to a Group Discussion, Novel Contributions, and Ignoring a Negative Stimulus. To investigate changes in externalizing behavior, specifically unsafe and defiant behavior, the following categories were used: Intentional Aggression Toward Staff, Repeated Noncompliance, Verbal Abuse, and Leaving the Activity Area. These categories were chosen based on the factor structure determined by Dunne (2020) who used confirmatory factor analysis to determine the utility of using composite scores of these behaviors to measure change over time. However, due to low estimates of internal reliability in the current study, these categories were modeled individually rather than as composites.

Social Responsiveness Scale, Second Edition

The Social Responsiveness Scale, Second Edition (SRS-2; Constantino & Gruber, 2012) is a norm-referenced parent-report measure that quantifies parent impressions of a child's social skills. Parents complete 65 items about their child's behavior using a Likert scale where a rating of 1 indicates, "not true," and a rating of 4 indicates, "almost always true" (Constantino & Gruber, 2012). Items correspond to five subscales: social awareness, social cognition, social communication, social motivation, and restricted interests and repetitive behavior (Constantino & Gruber, 2012). Importantly, for the present study items comprising the restricted interests and repetitive behavior subscale were excluded from the total score; while these items may be related

to social behavior, they were not the focus of the present study. Further, including items from this subscale would have likely had a strong relationship with child diagnosis status, which was the focus of the present study (i.e., restricted interests and repetitive behavior are core diagnostic features of ASD but not ADHD).

Higher SRS-2 scores suggest that the child presents with a greater degree of social impairment (Constantino & Gruber, 2012). Previous studies have demonstrated evidence of strong internal consistency in samples of children 4 to 17 years of age ($\alpha > 0.91$; Constantino & Gruber, 2012; Wigham et al., 2011), and evidence of moderate to strong convergent validity with other common measures such as the Social Communication Questionnaire ($r = 0.58 – 0.68$; Rutter, Bailey et al., 2003), Strengths and Difficulties Questionnaire ($r = 0.73$, $rs = 0.70$; Goodman, 1997), ADI-R ($r = 0.59$), and Checklist for Autism Spectrum Disorder ($r = 0.40$; Mayes, 2012) in samples of both young children and adolescents (Charman et al., 2007; Constantino & Gruber, 2012; Murray et al., 2011; Wigham et al., 2011).

According to Duvekot and colleagues (2014), the SRS-2 demonstrated evidence of strong ability to identify children with ASD comparable to the ADOS-2 (AUC = 0.59) and the Developmental, Dimensional, and Diagnostic Interview (AUC = 0.91; Skuse et al., 2004). Hus and colleagues (2013) demonstrated that among school-age children with ASD, parent-reported CBCL Externalizing scores strongly predicted and accounted for a significant portion of the variance in raw scores on the SRS-2 ($R^2 = 0.45$). In other words, higher scores on the SRS-2 have been associated with a greater degree of behavior problems as measured by the CBCL that were not necessarily specific to ASD (Hus et al., 2013). This pattern of findings was also significant in analyses of neurotypical siblings of children with ASD ($R^2 = 0.31$), suggesting that the SRS-2 may be a meaningful tool to assess social skills and externalizing problems even

among populations of typically-developing children (Hus et al., 2013). Also of note is the degree to which the presence of behavior problems can interfere with the ability of the SRS-2 to adequately measure social skills; children with typical social skills but high problem behavior had comparable raw scores to children with poor social skills and little problem behavior (Hus et al., 2013). Hus et al. (2013) recommended mitigating this interference by using the residuals of SRS-2 raw scores while controlling for externalizing behavior using CBCL scores. Other child-level factors that were associated with high scores on the SRS-2 included greater age, low language ability, and low nonverbal cognitive ability (Hus et al., 2013). Despite these limitations, the SRS-2 is a better instrument than many other rating scales to measure social outcomes associated with interventions for children with social skills problems due to its design for use with socially-impaired populations (Williams White et al., 2006).

Child Behavior Checklist

The CBCL is a parent- and teacher-completed rating scale that identifies impairment in internalizing and externalizing domains. Each domain has multiple syndrome scales (Achenbach & Rescorla, 2001). The Internalizing Domain scale includes scores from the Anxious/Depressed, Withdrawn/Depressed, and Somatic Complaints syndrome scales (Achenbach & Rescorla, 2001). The Externalizing Domain includes Rule Breaking Behavior and Aggressive Behavior syndrome scales (Achenbach & Rescorla, 2001). In addition, Social Problems, Thought Problems, and Attention Problems syndrome scales, as well as a Total Problems score, are available (Achenbach & Rescorla, 2001). Confirmatory factor analysis supported the use of this structure (Achenbach & Rescorla, 2001; Pandolfi et al., 2012). The respondent completes 113 items on a 3-point Likert scale wherein a rating of 0 indicates “Not True” and a rating of 2 indicates “Very True or Often True” (Achenbach & Rescorla, 2001). Raw scores can be

transformed into *T* or *Z* scores for the purposes of comparison with a normative sample (Achenbach & Rescorla, 2001).

The CBCL has demonstrated evidence of strong test-retest reliability (intraclass correlation coefficient = 0.95 – 1.00), internal consistency (α = 0.63 – 0.97), criterion validity (sensitivity = 80 – 88%), and construct validity (Achenbach & Rescorla, 2001). In addition, the CBCL has evidence of moderate to strong convergent validity with the Conners, Third Edition (*r* = 0.71 – 0.89; Conners, 2008) and the Behavior Assessment System for Children, Third Edition (*r* = 0.38 – 0.89; Reynolds & Kamphaus, 2015). A major advantage of the CBCL is its ability to provide both broad and narrow information across multiple domains, making it an efficient tool for screening and diagnostic assessment (Wetherbee & Achenbach, 2002). The CBCL has a long history of use for clinical purposes, including identifying children with oppositional defiant disorder who are at risk of developing mood disorders in adolescence (Masi et al., 2015). Among children with ASD, the CBCL demonstrated evidence of strong ability to discriminate between children with both ASD and an emotional/behavioral disability and children with ASD alone (Pandolfi et al., 2012). Further, subscales have been derived from the CBCL's items and structure to measure mania with evidence of excellent construct validity (comparative fit index = 0.97; Papachristou et al., 2016) and identify oppositional defiant disorder among children with an existing diagnosis of ADHD (area under the curve = 0.78 – 0.87; Biederman et al., 2008). Abbreviated parent-report forms also demonstrate evidence of internal consistency (α = 0.91), correlation with the long form (*r* = 0.95), and sensitivity to diagnoses of ADHD and ASD (Piper et al., 2014).

In addition to its clinical utility, the CBCL has a long history of measuring externalizing behavior and social skills across children and contexts for research purposes (Pecora et al.,

2002). In general, Wetherbee and Achenbach (2002) recommended that raw scores be used for statistical analyses as they tend to produce more accurate variances than transformed scores. However, if statistical analyses combine children of different genders or age groups and the population of interest is likely to achieve scores greater than the mean of the norm sample (e.g., a clinically-referred sample of children with oppositional defiant disorder), standardizing scores by transforming them into *T* scores is recommended (Wetherbee & Achenbach, 2002). If the population of interest is likely to represent a range of behavior, genders, and ages, converting raw scores to *Z* scores is recommended (Wetherbee & Achenbach, 2002). These transformations generally yield more accurate statistical analysis (Wetherbee & Achenbach, 2002). Other common statistical challenges when analyzing data from the CBCL include non-normal distributions and multicollinearity of subscales, both of which inflate Type II error (Wetherbee & Achenbach, 2002).

Statistical Analysis Plan

Preliminary Checks

Collinearity and Skew. Zero-order correlations among demographic variables, behavior frequencies, and parent rating scale scores were calculated to describe the data as well as to examine the possibility of multicollinearity; inspection of the correlations showed no predictor-predictor correlation to be problematic (i.e., there were no correlations greater than an absolute value of 0.90; Tabachnick & Fidell, 2012, p. 106). Of more concern was the potential for skew given the nature of the behavioral count data. Examination of histograms of both prosocial and externalizing scores revealed relatively normal distributions with some slight right skew.

Honeymoon Period. Previous research has suggested that the first week of data collected in the STP may be unreliable, as campers are becoming familiar with the point system and may

exhibit higher rates of desirable behavior and lower rates of challenging behavior during this period, sometimes referred to as a "honeymoon period" (e.g., Lopata et al., 2006; Pelham et al., 2000). To assess this in the data for the current study, daily averages for the first week of behavioral data were compared to daily averages from the second week. Although there were three significant differences between the first and second week (for Contributing to Group Discussion, $t(116) = 3.65$, $p < 0.001$; Novel Contribution, $t(116) = -7.79$, $p < 0.001$; and Ignoring a Negative Stimulus, $t(116) = -2.54$, $p = 0.012$), the means for two of these behaviors actually increased (Novel Contribution and Ignoring increased from 1.86 to 3.21 and 3.55 to 4.47 respectively), which is the opposite of what would be expected in a "honeymoon period." In other words, only one prosocial behavior decreased (Contributing to Group Discussion), and none of the externalizing behaviors exhibited a change from the first to second week (all ps > 0.05). Given these findings, data from the first week of camp were retained in all analyses, and the first day of camp was considered the baseline daily measurement.

Partial Absences. Based on examination of the first and last timestamps of each date for each camper, which was between the hours of 8:45 A.M. and 3:15 P.M., campers present for less than five hours of camp on any given day would be considered absent (i.e., data treated was missing) in order to ensure that their behavior counts were not predicated when given fewer opportunities within a day. An exception to this rule was made for dates in which at least one entire bunk and all associated campers were present less than five hours during the exact same time period; investigation of the timestamps for these dates showed that technology failures had prevented real-time data entry. Therefore, these data points were considered valid and retained (i.e., not missing). In total, partial absences resulted in treating 24 days as "missing" out of the 24 total days of the camp. These deletions impacted 20 participants; three of these had two days

deleted and one had three days deleted due to partially missing data. Importantly, all campers were retained for analysis; only these time points were treated as missing.

Multilevel Models

Multilevel models were conducted for each of the three prosocial and four externalizing behaviors separately using the *R* lme4 package, with Level 1 including a maximum of 19 measurements per child, Level 2 with $n = 117$ children, and Level 3 representing bunk membership (small group; $n = 10$). Time was a Level 1 predictor coded according to the number of days since the first measurement, starting at 0 days; as such the intercept represents the first measurement and the linear growth rate is in behaviors per day since baseline. Weekends, holidays, and field trip days when data were not collected were counted in the days since baseline, for a total of 31 calendar days.

Growth Model Functional Form. To evaluate research question 1, the optimal functional form of children's change over time for each behavior, prior to testing predictors, was evaluated by testing growth both as a linear function (i.e., no acceleration or deceleration) and as a quadratic function (i.e., allowing for one acceleration or deceleration), and with and without child-level random slopes (bunk-level slopes were constrained as fixed for model estimability). Model fit was then evaluated using the Bayesian Information Criterion (BIC) which favors parsimonious models, as well as the significance of the fixed effects growth terms. Further, both linear (i.e., additive) and Poisson (i.e., exponential) models were specified to ensure that linear model results were not an artifact of skew in the data; for this comparison, if the linear model results did not differ from the Poisson model results, the linear model was retained for clarity.

Predictors. To test the second and third research questions, predictors were added to the growth model. Specifically, diagnosis status (ADHD, ASD, and comorbid ADHD and ASD) was

coded into a set of two effect-coded Level 2 predictors to represent the three groups, with ADHD as the reference group. Parent-reported, pre-intervention child behaviors as measured by the SRS-2 and CBCL were also treated as Level 2, child-level, time-invariant predictors, standardized into *z*-scores for ease of results interpretation. Control variables were also added to tease out the unique effects of diagnosis and parent ratings, including standardized mean bunk age as a Level 3 predictor, child returner status (effect-coded with +1 = returner) as a Level 2 predictor, and child gender (effect coded with +1 = male) as another Level 2 predictor.

Chapter 4: Results

Descriptive Statistics

Zero-order correlations, along with descriptive statistics, among all predictors and first and last counts of each behavior category are provided in Table 3. Prosocial outcome variables correlated significantly with one another, as did some externalizing categories. Comorbid diagnosis significantly correlated with Novel Contributions. The CBCL and SRS-2 were also significantly correlated with diagnosis. The SRS-2 was significantly correlated with Novel Contributions at the final time point. Mean bunk age was significantly correlated with Novel Contributions at the final time point, ASD diagnosis, and comorbid diagnosis. Means and standard deviations of all measures by diagnostic group are presented in Table 4.

Intraclass Correlation Coefficients

The intraclass correlation coefficients (ICC) of bunk (Level 3) and child (Level 2) of all intercept-only models were calculated. In general, ICCs of bunk were low, ranging from 0 to 0.04 across models. In other words, between 0 and 4% of variance were accounted for by bunk membership across all models. ICCs at the child level ranged from 0.07 to 0.52. Across models, individual children accounted for between 7 and 52% of variance in the intercept-only models. ICCs of prosocial behavior categories were generally higher than externalizing categories. All ICC values are presented in Table 5.

Research Question 1: Growth Rate Model Results

As previously mentioned, optimal growth modeling form was evaluated using linear growth, quadratic growth, normal versus Poisson distributions, and with and without random slopes. First, it should be noted that a three-level model was retained rather than a two-level model to best reflect the structure of the data in the treatment context, even though bunk

membership explained little variance in child outcomes. Next, many of the attempted models, especially those specified with a Poisson distribution and those with random slopes, failed to converge on a solution. Across most models, except Contributing to Group Discussion and Novel Contribution, mean quadratic growth rates were not significant, indicating that linear growth better captured the data. For brevity and consistency given the model fit indices, for all outcomes, linear models with random intercepts and linear slopes were used for testing research questions 2 and 3. Additionally, for most models, the linear growth rate was constrained to be fixed across children; the exception was that for three outcomes (Intentional Aggression to Staff, Repeated Noncompliance, and Leaving the Activity Area), growth rates were allowed to vary across children as this was the better fitting model for those data.

The final growth model results across all measures are provided in Table 6. As can be seen, the average baseline (Day 1) prosocial behaviors across all campers, taking into account bunk membership, ranged from 2.49 to 5.38. For externalizing behaviors, average baseline counts ranged from 0.11 to 0.55. More interestingly, over the course of the camp experience, the mean daily change over time was significant for all prosocial behavior categories. There was a negative predicted slope for Contributing to Group Discussion which indicated that children decreased an average of 0.02 counts per day on this behavior during their camp participation, which amounted to a total decrease of a 0.06 count decrease since baseline ($p = 0.007$). Further, there were positive slopes for Novel Contribution with a predicted 0.07 count per day increase ($p < 0.001$) and Ignoring a Negative Stimulus with a predicted 0.14 count per day increase ($p < 0.001$), amounting to respective 2.10 and 4.20 count increases from baseline. Verbal Abuse was the only externalizing behavior category predicted to demonstrate significant change over the

course of camp, with a predicted 0.02 count per day increase equivalent to a predicted total 0.60 count increase from baseline to the final timepoint ($p < 0.001$).

Control Variable Relationships With Baseline Behaviors and Behavioral Growth Rates

Control variables of gender, returner status, and mean bunk age, as well as interaction terms with time, were added to control for group and child characteristics. Full results of growth models with only demographic predictors are presented in Model 1 in Tables 7-13. Patterns of significance of intercepts and growth rates remained the same as in the empty models. Broadly, returner status emerged as the most meaningful predictor across behaviors, showing that returners were 0.41 counts lower than average on baseline levels of Novel Contribution ($p = 0.047$) and 0.40 counts higher than average on baseline levels of Leaving the Activity Area ($p = 0.007$), holding all else constant. Clinically, this indicates a less desirable set of behaviors at baseline for returning campers. Mean bunk age accounted for a significant degree of variance in slope of Novel Contribution ($p = 0.003$); specifically, for every standard deviation increase in mean bunk age, there was a predicted 0.02 count increase per day in Novel Contributions holding all else constant. Gender also significantly predicted baseline ($p = 0.006$) and growth ($p < 0.001$) of Ignoring a Negative Stimulus. Male campers were expected to have -1.00 counts of Ignoring at baseline compared to average, or 2.00 counts fewer than female campers, holding all else constant. However, male campers were predicted to demonstrate a 0.07 count increase per day in Ignoring a Negative Stimulus compared to average, or a 0.14 count increase per day compared to female campers, all else held constant. Other than these exceptions, groups were predicted to be similar in baseline level and growth rate of behavior based on control variables.

Research Question 2: Differences Among Diagnoses

To address the relationship between behaviors and diagnosis, four variables were added to the models: ASD status, comorbid status, and interaction terms between diagnosis status and time. Diagnosis variables were entered into the models both without rating scales (see Model 2 in Tables 7-13), then with all predictors (see Model 4 in Tables 7-13).

Prosocial Behaviors and Diagnosis

Broadly, diagnosis significantly predicted frequency of Contributing to Group Discussion and Novel Contribution. Comorbid diagnosis significantly predicted Contributing to Group Discussion at baseline ($p = 0.034$); specifically, children with comorbid diagnoses were expected to enter camp with 0.97 fewer instances of Contributing to Group Discussion compared to average, all else held constant. The effect of comorbid diagnosis on Contributing to Group Discussion was no longer significant when entered into the model with SRS-2 scores. While ASD status was not a significant predictor of growth in the model of Contributing to Group Discussion with only diagnosis variables, it was significant when SRS-2 scores were entered into the model ($p = 0.039$). Specifically, ASD status significantly predicted a -0.02 count per day decrease compared to average in Contributing to Group Discussion over time holding all else constant. This relationship is graphed in Figure 2. Comorbid ($p = 0.002$) and ASD ($p = 0.012$) status both significantly predicted growth in Novel Contributions. Children with ASD were predicted to demonstrate a 0.02 count per day increase in Novel Contributions over time while children with comorbid diagnoses were expected to demonstrate a decrease of 0.02 counts per day, holding all else constant. When entered into the model with SRS-2 scores, only ASD status remained a significant predictor ($p = 0.032$), with a predicted increase of 0.02 counts per day when all else was held constant. This is graphed in Figure 3. Further, while returner status was a

significant predictor of Novel Contribution at baseline in Model 1, it was no longer significant after accounting for diagnosis predictors in Models 2 and 4. Diagnosis did not significantly predict Ignoring a Negative Stimulus at baseline or over time. Patterns of significance of control variables did not change after adding diagnosis to prosocial models. In summary, campers generally did not differ in prosocial behavior at baseline based on diagnosis, and ASD diagnosis appeared to be a stronger predictor of prosocial behavior over time than comorbid status, particularly when SRS-2 scores were entered into the model. Full results of prosocial models are presented in Models 2 and 4 in Tables 7-9.

Externalizing Behaviors and Diagnosis

Overall, diagnosis was not a strong predictor of externalizing behaviors. Comorbid diagnosis significantly predicted an increase of 0.56 counts of Intentional Aggression at baseline compared to average, holding all else constant ($p = 0.031$). Comorbid diagnosis also predicted an increase of 0.02 counts per day of Verbal Abuse after accounting for CBCL score in Model 4, holding all else constant ($p = 0.027$). This relationship is graphed in Figure 4. ASD ($p = 0.047$) and comorbid ($p = 0.047$) diagnosis both predicted baseline Leaving the Activity Area behavior before accounting for CBCL score; ASD status predicted a significant decrease of 0.38 counts while comorbid status predicted a significant increase of 0.42 counts in Leaving the Activity Area compared to average, holding all else constant. However, these diagnosis variables were no longer significant in Model 4 with all predictors. Diagnosis was not predictive of baseline levels or change in Repeated Noncompliance. Patterns of significance of control variables remained the same as in Model 1. In general, these results indicate that comorbid diagnosis was a stronger predictor of externalizing behavior than having a diagnosis of ASD alone, though both diagnoses

significantly contributed to only a minority of models. Full results of externalizing models are presented in Models 2 and 4 in Tables 10-13.

Research Question 3: Effect of Baseline Parent Rating Scales

To address the relationship between behavior and pre-camp rating scales without the influence of diagnosis, two variables were added to the models. For prosocial behaviors, baseline SRS-2 score and an interaction term between SRS-2 score and time were added. For externalizing behaviors, baseline CBCL score and an interaction term between CBCL score and time were added. Rating scale variables were entered into the models first without diagnosis (see Model 3 in Tables 7-13), then diagnosis was added (see Model 4 in Tables 7-13).

Prosocial Behaviors and Rating Scales

In the model of Contributing to Group Discussion, SRS-2 score was not a significant predictor; however, mean bunk age became a significant predictor of slope after adding SRS-2 scores to the model ($p = 0.023$). Specifically, for every standard deviation increase in mean bunk age there was a predicted 0.02 count increase in Contributing to Group Discussion at baseline with all else held constant. SRS-2 score was also not a significant predictor of Ignoring a Negative Stimulus at baseline or over time. However, SRS-2 score was significantly predictive of growth in Novel Contribution ($p = 0.002$); for every standard deviation increase in SRS-2 score, there was a predicted 0.02 count per day decrease in Novel Contributions compared to average, holding all else constant. Because the predicted slope was negative, campers with higher SRS-2 scores (i.e., poorer parent-reported social skills) at baseline were predicted to show a significant decrease in Novel Contribution over the course of camp. This relationship is graphed in Figure 5. Other than mean bunk age becoming significant in the model of Contributing to Group Discussion, patterns of significance of control variables remained the

same as in other models. Overall, SRS-2 score appeared to be a largely nonsignificant predictor of prosocial behavior.

Externalizing Behavior and Rating Scales

CBCL score predicted baseline Repeated Noncompliance frequency before accounting for diagnosis ($p = 0.039$); for every standard deviation increase in CBCL score, there was a predicted 0.06 count increase in Repeated Noncompliance at baseline. However, CBCL score was nonsignificant in the full model of Repeated Noncompliance that included all predictors. Further, while mean bunk age was a significant predictor of Repeated Noncompliance in the model of only rating scales suggesting that for every standard deviation increase in mean bunk age there was a predicted 0.06 count decrease in Repeated Noncompliance at baseline holding all else constant ($p = 0.046$), it was no longer a significant predictor after adding diagnosis in Model 4. CBCL score significantly predicted baseline Verbal Abuse before ($p = 0.008$) and after ($p = 0.015$) accounting for diagnosis. Similarly, CBCL score predicted baseline Leaving the Activity Area behavior before ($p = 0.005$) and after ($p = 0.021$) accounting for diagnosis. In Model 3 without diagnosis, for every standard deviation increase in CBCL score there was a predicted 0.42 count increase in Leaving the Activity Area at baseline holding all else constant. In Model 4 with diagnosis, each standard deviation increase in CBCL score predicted a 0.36 count increase in Leaving the Activity Area at baseline holding all else constant. The CBCL was not significantly predictive of Intentional Aggression or growth over time in any category. Other than mean bunk age in the model of Repeated Noncompliance, significance of control variables remained the same as in prior models. In summary, the CBCL was a relatively consistent predictor of externalizing behavior at baseline across externalizing behaviors, though not a significant predictor of growth.

Chapter 5: Discussion

The purpose of this study was to examine social and behavioral growth during Apex Summer Camp, which is a modified version of the STP that includes children with ADHD, children with ASD, and children with comorbid ADHD and ASD. The study constructed growth models for seven behavior categories and analyzed the predictive relationship of diagnosis and parent rating scales using hierarchical linear modeling. Results indicated that, on average, frequency of prosocial behavior tended to increase over time while most externalizing behaviors did not change significantly. Comorbid diagnoses of ASD and ADHD predicted less growth than average in some prosocial behaviors as well as higher levels of some externalizing behaviors at baseline, and ASD diagnosis significantly predicted change in one out of three prosocial behaviors. Finally, SRS-2 scores at baseline only predicted growth in one prosocial category; CBCL scores predicted baseline levels of some externalizing behavior, although not change over time. These results are explored in more detail below.

Main Findings

Research Question 1: Growth Rate Models

Several main findings emerged from the results of the growth models of prosocial data. First, campers at Apex demonstrated a significant increase in frequency of Novel Contribution and Ignoring a Negative Stimulus behaviors over time; the effect size of both categories was considered moderate to large. Campers exhibited a significant decrease in Contributing to Group Discussion frequency. The finding that Novel Contributions increased while Contributing to Group Discussion decreased aligns with a common clinical goal at Apex; while many campers can easily repeat rules, which is the most common way to earn points for Contributing, it is more challenging to provide a new idea to the discussion by responding to an open-ended question or

elaborating on someone else's idea, which is the criterion to earn points for a Novel Contribution. To encourage Novel Contributions, campers' individualized goals often changed over time to increase the frequency of Novel Contributions over Contributing to Group Discussion. The opposite growth pattern of these two variables may validate that Novel Contributions replaced instances of Contributing for many campers. These results indicate that Apex may be a clinically meaningful intervention to support the growth of social skills. This finding is consistent with prior research on the STP that has demonstrated improvements in peer relationships and social skills during and after the STP (Chronis et al., 2004; Pelham et al., 2000; Pelham & Fabiano, 2008). Results of the current study are especially important in light of limited existing research to support Apex as a modified STP treatment package; significant growth in prosocial behavior while participating in Apex further supports the legitimacy of Apex as an intervention.

Regarding models of externalizing behaviors, there was no significant change in most externalizing behavior categories over time, except for a statistically significant predicted increase in Verbal Abuse. Broadly, this suggests that children who participate in Apex may not show improvement in externalizing behaviors. While potentially disheartening from a clinical perspective, this finding is consistent with Dunne (2020) who demonstrated a statistically significant increase in defiant behaviors (i.e., Verbal Abuse, Repeated Noncompliance, and Leaving the Activity Area) over time in the Apex model. Inconsistent with this result are findings from Chronis et al. (2004) and Pelham and Fabiano (2008) who documented decreases in externalizing behaviors among participants in the STP. One reason for these mixed findings across the literature could be that Apex is an adaptation of the STP with some meaningful differences, including a lower dose of intervention. Another possible reason for these discrepant

findings is the inclusion of children with ASD and comorbid ASD and ADHD in this sample; these diagnostic variables are discussed in more detail below.

Of the control variables, mean bunk age significantly predicted Novel Contribution behaviors when all predictors were entered into the models; older campers demonstrated a significantly higher growth rate of Novel Contributions than younger campers. This is unsurprising considering that older children are more likely to have more developed communication skills than younger children and may also be given more opportunities to expand on ideas compared to younger children who may need more direct instruction, particularly early in camp as they learn rules and routines. Gender was a significant predictor of Ignoring a Negative Stimulus, both at baseline and over time; while male campers were likely to start camp with a lower frequency of Ignoring compared to average, they demonstrated a strong upward trajectory throughout camp. It is possible that this gender difference is due to differences in outward presentation of neurodevelopmental disorders between girls and boys; boys may be more likely to be perceived as impulsive and likely react to negative stimuli, while girls may be perceived as more withdrawn or inattentive (e.g., Rucklidge, 2010).

The lack of significance of returner status in prosocial models is worthy of further examination. Surprisingly, returner status only predicted Novel Contributions at baseline, with returning campers expected to demonstrate a significantly lower rate of Novel Contributions than new campers at the start of camp. It may be expected that returning campers would demonstrate higher frequencies of prosocial behavior at baseline compared to new campers given their previous experience with the intervention model; however, this was not borne out in the results of the present study. It may be that campers who returned to Apex demonstrated a continued need for social skills intervention while campers who did not return developed sufficient social

skills to be successful in traditional, community summer activities, resulting in returning campers presenting similarly to new campers. In any case, understanding which campers return and why represents one important direction for future research; determining an optimal dosage of Apex may help families and clinicians make decisions about maximum possible benefit from the intervention for individual children.

In the models of externalizing behaviors, gender, returner status, and mean bunk age largely did not predict externalizing behavior at the start of camp or over time, excepting the result that returning campers were expected to have higher rates of Leaving the Activity Area. This is somewhat consistent with Dunne (2020) who found that returner status predicted baseline levels of defiant behavior which included Leaving the Activity Area, but not unsafe behavior. The finding that gender, returner status, and mean bunk age did not significantly predict levels of most externalizing behavior at baseline or over time indicates that campers came to Apex with comparable frequencies of externalizing behavior, regardless of these characteristics. Clinically, this may mean that campers, even those that have attended Apex in prior years, are unlikely to present differently in their frequency of externalizing behavior and, as a result, indicates that differentiating campers based on these characteristics is unnecessary.

Research Question 2: Differences Among Diagnoses

Prosocial Behavior and Diagnosis. Regarding the relationship between diagnosis and prosocial behaviors, comorbid diagnosis predicted baseline levels of Contributing to Group Discussion before adding SRS-2 score to the model. ASD and comorbid diagnosis status significantly predicted frequency of Novel Contributions over time, although not at baseline, and only ASD remained significant after taking into account SRS-2 score. This finding implies that campers with ADHD, campers with ASD, and campers with comorbid ASD and ADHD entered

camp with a comparable frequency of prosocial behaviors, and that changes in prosocial behavior were slightly different across diagnostic groups. Specifically, after accounting for baseline SRS-2 scores, campers with ASD were expected to demonstrate significantly more improvement than average in their Novel Contributions. No variable related to diagnosis status predicted Ignoring a Negative Stimulus. This result adds to the small body of existing literature that suggests that STP models are applicable to children with ASD as well as children with ADHD to support social skill development (Dunne, 2020; Mrug & Hodgens, 2008; Wymbs et al., 2005). More broadly, it supports the conclusion that it may be beneficial to combine these diagnostic groups in intervention settings that make use of evidence-based practices for both children with ASD and children with ADHD (e.g., Antshel et al., 2016). Little known research to date has examined prosocial growth in children with ASD, children with ADHD, and children with comorbid ASD and ADHD who all participated in the same intervention; results from the present study provide evidence that separating children for social skills intervention based on diagnosis alone may be unnecessary and inefficient, particularly for intensive programs such as the STP and Apex.

It is notable that children with comorbid diagnoses demonstrated a different trajectory in Novel Contribution behavior from other diagnostic groups when only diagnosis and control variables were entered into the model. It might be expected that children with comorbid ASD and ADHD had more severe symptomatology, and therefore more impaired social skills and resistance to treatment (Antshel et al., 2011; Factor et al., 2017; Rao & Landa, 2014). With respect to Novel Contributions, this was borne out in the present study. However, comorbid diagnosis did not predict growth over time in Contributing to Group Discussion or Ignoring a Negative Stimulus compared to average, and it was no longer significant in the model of Novel

Contributions after controlling for SRS-2 scores. Thus, children with comorbid diagnoses of ASD and ADHD may still stand to benefit from interventions like the STP in terms of growth in prosocial behavior. In fact, models predicted that campers with comorbid diagnoses still showed upward trajectories in Contributing to Group Discussion and Ignoring a Negative Stimulus, even if the growth was statistically nonsignificant or slower than campers who had ASD or ADHD alone; campers with comorbid diagnoses showed an increase in Contributing to Group Discussion even while other diagnostic groups were predicted to decrease. Few studies to date have examined social skills interventions for children with comorbid ASD and ADHD (e.g., Davis & Kollins, 2012; Joshi et al., 2017). Only one study on the STP, specifically of the Apex model, has examined the prosocial trajectory of school-age children with comorbid diagnoses and determined that comorbid status largely did not differentiate campers from their peers with one diagnosis (Dunne, 2020). The results of the present study provide further insight into social skills interventions for children with comorbid ASD and ADHD and suggest that participants in programs like Apex who have comorbid diagnoses may make meaningful improvements in social skills. Further research may help to determine the usefulness of differentiating treatment goals and strategies to bolster intervention effectiveness for campers with comorbid diagnoses.

Externalizing Behavior and Diagnosis. With respect to externalizing behavior and diagnosis, comorbid diagnosis predicted a higher level of baseline Intentional Aggression. Both ASD diagnosis and comorbid diagnosis status predicted baseline Leaving the Activity Area frequency, with children with ASD expected to have a lower rate than average and children with comorbid diagnoses expected to have a higher rate than average. These variables were no longer significant after CBCL score was entered into the model. Finally, comorbid status predicted a slight but significant increase in Verbal Abuse over time only after controlling for CBCL score.

These results indicate that children with comorbid ASD and ADHD may enter camp with a higher degree of some externalizing behaviors and may show poorer trajectories in some externalizing behaviors compared to average. This finding is somewhat consistent with Dunne (2020) who found that diagnosis predicted defiant behavior (i.e., Verbal Abuse, Leaving the Activity Area, Repeated Noncompliance). Broader literature has suggested that children with comorbid presentations often experience more intense symptomatology than children with ASD or ADHD alone (Antshel et al., 2011; Factor et al., 2017; Rao & Landa, 2014); therefore, it is expected that campers with comorbid diagnoses might begin treatment with higher levels of externalizing behavior, in this case Intentional Aggression to Staff, and show more resistance to intervention over time, as evidenced by the model of Verbal Abuse (Reierson & Todd, 2008). However, it is important to note that, out of sixteen terms for comorbid diagnosis across four sets of models, only four were statistically significant predictors of externalizing behavior, meaning that campers with comorbid ASD and ADHD were consistent with average frequencies more often than not. Further, ASD diagnosis only significantly predicted baseline Leaving the Activity Area in only one model without CBCL score; otherwise, ASD did not predict baseline or growth of behavior in any of the other models.

These findings may indicate that diagnosis does not distinguish campers' behaviors at baseline or over time in many cases. Implications of such a finding could further validate that Apex is applicable across diagnostic groups. At a minimum, these findings fail to provide evidence that children with ASD and comorbid ASD and ADHD differ significantly from children with ADHD in an STP model in terms of externalizing behavior. In the context of broader literature, this is consistent with previous studies that have demonstrated behavioral benefit associated with the STP for campers with ASD alone (Mitchell et al., 2015; Mrug &

Hodgens, 2008; Wymbs et al., 2005) and contributes to the limited existing literature regarding possible behavior intervention models for children with comorbid ASD and ADHD (Antshel et al., 2011; Joshi et al., 2017).

Question 3: Effect of Baseline Parent Rating Scales

Only growth in frequency of Novel Contributions over time was predicted by baseline SRS-2 score. Baseline levels and growth of Contributing to Group Discussion or Ignoring a Negative Stimulus behaviors were not predicted by parent report of social skills prior to camp as measured by the SRS-2. Making Novel Contributions requires a relatively sophisticated set of social-communication skills, so it is unsurprising that growth was predicted by SRS-2 scores; children with high SRS-2 scores, and therefore more parent-reported social impairment at baseline, had less growth in the frequency of Novel Contribution behavior than children with low SRS-2 scores. There are several possible explanations for the lack of statistically significant relationship between the other prosocial variables and SRS-2 scores, the first of which is that there is not a predictive relationship between parent report of social skills and some prosocial behaviors. It is likely that there are naturally-occurring differences in the presentation of social behaviors at camp and at home; some variation in behavior across settings and contexts is to be expected by researchers and clinicians (Konold et al., 2019). The lack of relationship documented here could be a manifestation of this difference. In the same vein, parent report is inherently subjective to the parents' experience of children's behavior and is therefore unlikely to perfectly match the observations or perceptions of a clinician outside of the family.

Another possible explanation for the lack of relationship may be the methods used to measure social skills. The SRS-2 is a measure designed for children with ASD; while the restricted, repetitive behavior subscale was excluded in an effort to more accurately capture

social skills across diagnostic groups, it may not represent the best generalized measure of social skills across children with various neurodevelopmental disorder diagnoses. There may be some evidence to support this idea in the changes in significance across models with and without diagnosis, as well as the significant correlation between diagnosis and SRS-2 score. In addition, scores on the SRS-2 are vulnerable to interference from externalizing behaviors (Hus et al., 2013); it may be difficult for a parent to accurately report on their child's social skills if they are also exhibiting high rates of externalizing behaviors. It is also possible that these different methods of capturing social skills are simply fulfilling their unique purposes, as suggested by Volpe and Gadow (2010); parent rating scales may be beneficial for measuring development more broadly while direct behavior observations may be more likely to capture specific treatment targets that are difficult to assess with rating scales.

In any case, the lack of relationship between rating scales and observed prosocial behavior in this context raises questions about the utility of collecting parent rating scales focused on social skills. If there truly is no relationship between social skills observed by clinicians and parent rating scale scores, there is little apparent benefit to collecting parent rating scales, at least for the purpose of evaluating a statistical relationship. However, this study did not include a post-treatment measurement using parent rating scales which may show significant improvements in social skills compared to parent report at baseline. Examining parent report at multiple time points in conjunction with behavioral data may allow researchers to better determine the characteristics and causes of the discrepancy between behavioral data and parent report of social skills. The present study highlights the limitations of attempting to make sense of two different measurement methods to evaluate the same construct. More research is needed to inform the process of decision-making across methods and raters in the context of measuring

intervention outcomes and related predictors. While there may still be clinical benefit to including parent rating scales as a measurement tool, as discussed in more detail below, parent report of social skills appeared to have limited value for predicting prosocial outcomes among campers at Apex.

Regarding the effect of CBCL Externalizing scores, parent-reported CBCL scores at baseline predicted most externalizing behavior categories at baseline, but not over time. In other words, campers demonstrated significant differences in externalizing behavior counts based on parent report of externalizing behavior at the start of camp. This finding suggests that the externalizing composite of the CBCL may be a helpful measure of where campers start with externalizing behavior, but not necessarily how they develop over time. Few studies to date have used parent rating scales to predict externalizing behavior in the STP. Dunne (2020) used the CBCL Attention Problems subscale as a predictor for unsafe and defiant behavior (i.e., two composites that included the behavior categories used in the present study) and found that it did not predict these behaviors at baseline or over time in the Apex model. While not a direct comparison to the present study, Attention Problems scores may correlate with Externalizing scores and result in a similar pattern of findings, as there is a moderate correlation of 0.59 between Attention Problems and Externalizing subscales among the normative sample of the CBCL (Achenbach & Rescorla, 2001). Thus, the CBCL's prediction of baseline levels but not growth of externalizing behavior is somewhat unexpected in the context of the current literature. Notably, changes in patterns of significance in models with and without diagnosis may suggest an unexplored interaction between CBCL score and diagnosis. Given the significant correlation between diagnosis and rating scale scores, this is unsurprising. Future research may wish to further explore the nature of this relationship. Clinically, these results support the use of the

CBCL as a tool to determine which campers may be most likely to demonstrate defiant behaviors (i.e., Verbal Abuse, Repeated Noncompliance, and Leaving the Activity Area) early on in camp.

Limitations and Future Directions

Several limitations based on the statistical methods used in the present study warrant discussion. First, in general, externalizing behavior categories had low frequencies which presented a barrier to accurate modeling. Second, although quadratic terms for time fit well to some models, they were not used for the sake of consistency across categories; this means that analysis may not have captured behaviors with a U-shaped curve (i.e., those that increase and then decrease over time). Third, these behavior categories had poor internal reliability when combined into broad prosocial and externalizing categories, suggesting that they may not collectively measure two distinct constructs. This may be due to a real lack of cohesion, or could be due to "noise" in the data that may be an artifact of the sample; while children who attend Apex often have similar symptom profiles, children with ASD are likely to vary widely across all areas of development, and children with ASD and children with ADHD are assumed to necessarily be different from one another, although the nature and degree of this difference is still contested (e.g., van der Meer et al., 2012; Rommelse et al., 2018). Future research may continue to investigate the utility of using behavioral composites, and characteristics of participants that may significantly differentiate them from one another.

There are several opportunities for future research in study design as well. Within the STP model, several behavior categories may be worthy of further examination that were excluded here. For example, the prosocial categories used in this study are biased toward verbal communication abilities and are often adult-directed. The externalizing categories used here are also largely adult-directed and do not account for peer-directed behavior that may damage

relationships, such as teasing. Future studies may wish to further explore other behavior categories and continue to examine the relationships between the STP categories. It is possible that behavior categories that were not used in the present study demonstrate a stronger predictive relationship with rating scales or better internal reliability to derive composite scores for statistical and clinical use.

Considering external validity, characteristics of children in this sample are unlikely to be representative of the larger population of children with ADHD or children with ASD, which represents a substantial threat to external validity. Children who were enrolled in Apex were accepted into the program based on the decision of the clinical team. Children who were accepted relied primarily on verbal communication and had already acquired adequate adaptive skills to independently address physical needs such as toileting and eating. While findings from this study may be generalizable to children with similar symptom presentations and behavior profiles, they may not be applicable to many children with ADHD and children with ASD without replication among a broader range of clinical presentations (Shadish et al., 2002). However, consistent with empirically-supported practices for children with ADHD and children with ASD, the STP was designed as a naturalistic intervention to increase the likelihood of treatment gains also appearing in settings other than camp. Although this has yet to be investigated among children who participate in Apex specifically, this is an important feature that may bolster external validity.

Because the current study is not a randomized control trial, results of the study focus only on growth during Apex rather than making causal claims. As a result of this design, threats to internal validity are present and also warrant discussion. History is likely to be a salient threat to internal validity (Shadish et al., 2002). Other than attendance at Apex in prior years, detailed

information about previous treatment experiences or concurrent treatments outside of camp such as speech therapy was not collected. Future studies may wish to investigate differences among campers based on concurrent or previous intervention experiences. Regression to the mean is another possible threat to internal validity (Shadish et al., 2002). This may be in part due to random error in measurement, but may also be due to selection of participants. Most children who attended Apex did so because they had significant challenges with social and externalizing behavior. If possible, a randomized control trial of Apex represents a rigorous future direction for intervention research and would address these threats to internal validity.

Implications for Practice

Implications for the STP and Apex

Several clinical lessons can be drawn from the present study. Inherent in the results is the implication that participants in Apex do not demonstrate improvement in externalizing behaviors. In fact, there was a small but statistically significant increase in one externalizing behavior, Verbal Abuse, across the sample. This may suggest that the strategies used to target externalizing behaviors at Apex warrant reexamination. It may also signal a difference in the Apex sample from much of the other published work on the STP. Campers with comorbid ASD and ADHD seemed to have somewhat worse trajectories than average in this study, and these campers would have been excluded from other work that has focused solely on samples of children with ADHD or children with ASD. It is notable, though, that frequencies of externalizing behaviors were not only low across the sample compared to prosocial behaviors, but also variable across campers. Some campers had no instances of any of the included behavior categories for the duration of camp, while some campers were reported to have extremely high frequencies, such as one who had 94 recorded instances of Intentional Aggression in one day.

While this study fails to establish desirable growth for externalizing behavior while participating in Apex, it provides some valuable insights regarding the behavioral trajectories of children in different diagnostic groups. This is the second known study to examine school-age campers with comorbid ASD and ADHD who participated in an STP. Results suggest that campers with comorbid diagnoses may be more resistant to significant changes in prosocial and externalizing behavior and may show a different pattern of development in some cases, such as exhibiting a decrease in Novel Contribution while the rest of the sample demonstrated an increase that theoretically replaced Contributing to Group Discussion. In addition, campers with ASD showed a significantly stronger upward trajectory than average in Novel Contributions and were not significantly different in their frequencies of externalizing behaviors over time. The implication of these findings is that campers with comorbid diagnoses may warrant special attention and possible differentiation in their intervention experience; more research is needed to determine exactly what modifications may enhance outcomes associated with Apex for this population. A second implication is that campers with ASD are likely a good fit for Apex and STP interventions; continued inclusion of campers with ASD is supported by this study.

Lastly, the present study is among the first to examine the relationship between parent rating scales and the frequency of behaviors in an STP intervention. Scores on the CBCL seemed to be a strong predictor of some externalizing behavior, but the SRS-2 was a poor predictor of all prosocial behavior except for Novel Contribution. This may be because prosocial behaviors are more challenging for parents to report on as they may not often observe their child in a group of peers, as they are at camp. On the other hand, the externalizing behaviors included in this study are likely to be adult-directed, clearly discernable, and of high clinical concern for many parents. While it is possible that these results are due to measurement properties or a true mismatch

between parent report and observed behavior in some cases, it is also possible that the patterns of significance indicate a qualitative difference between prosocial and externalizing behaviors. In any case, results of the present study call into question the clinical utility of parent rating scales as a predictive measurement tool for campers at Apex. Rating scales may have importance when examining generalizability or parent perceptions of Apex, but results presented here suggest that camp staff are unlikely to gain insight into most categories of campers' behavior based on rating scale scores.

Implications for School-Based Practice

While findings from the present study are most applicable to Apex and similar intensive behavioral interventions, there are several implications for school-based practice. First and foremost, findings suggest that children with ASD and children with ADHD may both experience benefit while participating in the same interventions. While research literature indicates that children are often separated for intervention based on diagnosis, this may be unnecessary; intervention packages that capitalize on best practices for both children with ASD and children with ADHD may be effective, particularly in teaching social skills. Some differentiation may be warranted for children with comorbid ASD and ADHD who demonstrated a different trajectory of some behaviors. However, among campers who attended Apex, children with comorbid diagnoses demonstrated an increase in prosocial behaviors, even though their growth appeared slower than other groups, suggesting that some benefit may still be possible.

In addition to the conclusion that diagnosis alone does not dictate behavioral outcomes of intervention, the results of the present study suggest that social skills instruction may be associated with increases in prosocial behavior for children with ADHD. As mentioned previously, research on SST among samples of children with ADHD is mixed and has often

taken place in one-on-one or office settings. Findings presented here lend support to the legitimacy of SST as a possible intervention strategy to support development of social skills for children with ADHD within a comprehensive intervention package. Additionally, using SST in a naturalistic and structured setting may increase the generalization of social skills. For school-based practitioners, this may mean that designing social skills instruction for group settings and preparing other complementary strategies to support children with neurodevelopmental disabilities such as behavioral contingencies may increase the likelihood of significant changes in social skills.

Finally, among populations of students with neurodevelopmental disabilities there may be a limited relationship between what parents report on rating scales and what specific behaviors are observed by teachers and other school staff. There are a number of possible reasons for this lack of relationship discussed above that remain relevant to school-based practice. However, the considerations of using parent rating scales as a measurement tool may be different in schools. In some cases, completion of parent rating scales may be the only opportunity for a caregiver to contribute to a psychoeducational evaluation. Additionally, when both parents and teachers complete complementary versions of the same scale, it allows for direct comparison across raters and environments to make decisions about necessary supports for the child being evaluated and results in a more complete evaluation (Tripp et al., 2006). Not only do rating scales offer the opportunity to compare raters, but they can also compare students to peers in their age group or grade. In some cases, rating scales may be strongly recommended or required to better examine factors that may be contributing to a presenting problem. For example, for a student with a diagnosis of ADHD, a school psychologist may choose to administer a Conners, Third Edition (Conners, 2008) or BRIEF-2 (Gioia et al., 2015) to understand their executive function in a way

that may be challenging to assess through direct observation. All of these points underscore the importance of the use of parent rating scales in school settings, even if they do not directly predict many behaviors. School-based practitioners should remain aware, however, that rating scales may be a poor tool to measure progress frequently and may not strongly correlate with specific target behaviors or goals within a student's intervention program.

Conclusion

The present study explored prosocial and externalizing growth among children with ASD, children with ADHD, and children with comorbid ASD and ADHD who participated in Apex Summer Camp, a modified STP program. Results indicated that, while participants demonstrated overall increases in prosocial behavior, these changes were not consistently predicted by diagnosis or SRS-2 scores, particularly when both sets of predictors were in the same model. With respect to externalizing behaviors, participants in Apex were not predicted to demonstrate significant changes, except for an increase in Verbal Abuse. Comorbid diagnosis differentiated baseline levels of behavior and growth in a minority of models, indicating that campers with comorbid diagnoses began camp with higher levels of externalizing behavior than average in some categories. The CBCL Externalizing composite was a more consistent predictor of externalizing behavior frequency at baseline than the SRS-2 was for prosocial behavior categories.

The present study is the first known study to examine the relationship of parent report of social skills and externalizing behavior to observed frequencies of these behaviors in the context of an STP program. It is also among the first to construct growth models of individual behavior categories and account for the effects of diagnosis. While Apex is modeled on the STP, more research is needed to validate it as a stand-alone intervention program in its own right. Results

presented here suggest that there are possible benefits in the development of social skills for participants in Apex; ideally, causality should be established in the future through a randomized control trial, perhaps in a waitlist control design. Future research may also further investigate outcomes of children with comorbid neurodevelopmental disabilities. Results of the present study are consistent with prior research that has indicated more profound impairment in behavior among children with comorbid ASD and ADHD (Antshel et al., 2011; Factor et al., 2017; Rao & Landa, 2014; Reierson & Todd, 2018). More research is needed about what components of interventions enhance or dampen treatment response for the comorbid population, as well as more clearly defining the diagnostic differences between children with ASD, children with ADHD, and children with both disorders.

Prevalence of neurodevelopmental disorders has been increasing in recent years (e.g., Xu et al., 2018), and with it our knowledge of how to support this population of children. ASD and ADHD often impact individuals throughout their lives, leading to long-lasting effects for their families and communities (APA, 2013; Sasser et al., 2016). Evidence-based intervention programs can support these children as they develop social skills, emotional regulation, coping skills, and functional communication strategies (e.g., Fabiano et al., 2014; Wong et al., 2015). Programs such as Apex and the STP represent one bridge to meaningful participation in community settings for children and families impacted by ASD and ADHD; conducting intensive, evidence-based interventions in a naturalistic context may result in behavioral change in other environments. It is of great importance that interventions like Apex continue to be investigated to enhance the quality of life for families impacted by ASD and ADHD. With continued support and research, intensive interventions like Apex can continue to impact the lives of children with neurodevelopmental disabilities, setting them up for success as they grow.

References

Abikoff, H. (2009). ADHD psychosocial treatments: Generalization reconsidered. *Journal of Attention Disorders, 13*(3), 207-210. http://doi.org/10.1177/1087054709333385

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles.* University of Vermont.

Althoff, R. R., Ayer, L. A., Rettew, D. C., & Hudziak, J. J. (2010). Assessment of dysregulated children using the child behavior checklist. *Psychological Assessment, 22*(3), 609-617. http://doi.org/10.1037/a0019699

American Academy of Child and Adolescent Psychiatry. (2007). Practice parameter for the assessment and treatment of children and adolescents with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*(7), 894-921). http://doi.org/10.1097/chi.0b013e318054e724

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596.

Antshel, K. M., & Russo, N. (2019). Autism spectrum disorders and ADHD: Overlapping phenomenology, diagnostic issues, and treatment considerations. *Current Psychiatry Reports, 21*(5), 1-11. http://doi.org/10.1007/s11920-019-1020-5

Antshel, K. M., Polacek, C., Mcmahon, M., Dygert, K., Spenceley, L., Dygert, L., Miller, L., & Faisal, F. (2011). Comorbid ADHD and anxiety affect social skills group intervention treatment efficacy in children with autism spectrum disorders. *Journal of Developmental and Behavioral Pediatrics, 32*(6), 1-11. http://doi.org/10.1097/DBP.0b013e318222355d

Antshel, K. M., Zhang-James, Y., Wagner, K. E., Ledesma, A., & Faraone, S. V. (2016). An update on the comorbidity of ADHD and ASD: A focus on clinical management. *Expert*

Review of Neurotherapeutics, 16(3), 279-293. http://doi.org/10.1586/14737175.2016.1146591

Ashwood, K., Tye, C., Azadi, B., Cartwright, S., Asherson, P., & Bolton, P. (2015). Brief report: Adaptive functioning in children with ASD, ADHD and ASD + ADHD. *Journal of Autism and Developmental Disorders, 45*(7), 2235-2242. http://doi.org/10.1007/s10803-014-2352-y

Bagwell, C. L., Molina, B. S. G., Pelham, W. E., & Hoza, B. (2001). Attention-deficit hyperactivity disorder and problems in peer relations: Predictions from childhood to adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(11), 1285-1292. http://doi.org/10.1097/00004583-200111000-00008

Bailey, J. S., & Burch, M. R. (2018). *Research methods in applied behavior analysis* (2nd ed.). Routledge.

Baldwin, S., & Costley, D. (2015). The experiences and needs of female adults with high-functioning autism spectrum disorder. *Autism: The International Journal of Research and Practice, 20*(4), 483-495. http://doi.org/10.1177/1362361315590805

Bargiela, S., Steward, R., & Mandy, W. (2016). The experiences of late-diagnosed women with autism spectrum conditions: An investigation of the female autism phenotype. *Journal of Autism and Developmental Disorders, 46*(10), 3281-3294. http://doi.org/10.1007/s10803-016-2872-8

Barkley, R. A. (2015). Psychological assessment of children with ADHD. In R. A. Barkley (Ed.), *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (4th ed., pp. 455-474). Guilford Press.

Barkley, R. A. (2016). Recent longitudinal studies of childhood attention-deficit/hyperactivity disorder: Important themes and questions for further research. *Journal of Abnormal Psychology, 125*(2), 248-255. http://doi.org/10.1037/abn0000125

Barkley, R. A., & Edelbrock, C. (1987). Assessing situational variation in children's problem behaviors: The home and school situations questionnaires. In R. Prinz (Ed.), *Advances in behavioral assessment of children and families.* JAI Press.

Bauminger, N., Solomon, M., & Rogers, S. J. (2010). Externalizing and internalizing behaviors in ASD. *Autism Research, 3*(3), 101-112. http://doi.org/10.1002/aur.131

Bearss, K., Lecavalier, L., Minshawi, N., Johnson, C., Smith, T., Handen, B., Sukhodolsky, D., Aman, M., Sweizy, N., Butter, E., & Scahill, L. (2013). Toward an exportable parent training program for disruptive behaviors in autism spectrum disorder. *Neuropsychiatry, 3*(2), 169-180. http://doi.org/10.2217/npy.13.14

Berry, W., & Feldman, S. (1985). *Multiple regression in practice* (50th ed.). Sage. doi:10.4135/9781412985208

Biederman, J., Ball, S. W., Monuteaux, M. C., Kaiser, R., & Faraone, S. V. (2008). CBCL clinical scales discriminate ADHD youth with structured-interview derived diagnosis of oppositional defiant disorder (ODD). *Journal of Attention Disorders, 12*(1), 76-82. doi:10.1177/1087054707299404

Bijlenga, D., Tjon-Ka-Jie, J., Schuijers, F., & Kooij, J. (2017). Atypical sensory profiles as core features of adult ADHD, irrespective of autistic symptoms. *European Psychiatry, 43*, 51-57. http://doi.org/10.1016/j.eurpsy.2017.02.481

Bird, H. R., Shaffer, D., Fisher, P., & Gould, M. S. (1993). The Columbia impairment scale (CIS): Pilot findings on a measure of global impairment for children and adolescents.

International Journal of Methods in Psychiatric Research, 3(3), 167–176. https://psycnet.apa.org/record/1994-08032-001

Bishop, D. (2006). *Children's communication checklist 2nd ed.* Pearson.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 1-18. http://doi.org/10.3389/fpubh.2018.00149

Boonen, H., Maljaars, J., Lambrechts, G., Zink, I., Van Leeuwen, K., & Noens, I. (2014). Behavior problems among school-aged children with autism spectrum disorder: Associations with children's communication difficulties and parenting behaviors. *Research in Autism Spectrum Disorders, 8*(6), 716-725. http://doi.org/10.1016/j.rasd.2014.03.008

Brown, T. E. (2018). *Brown executive function/attention scales.* Pearson.

Buie, T., Campbell, D. B., Fuchs, G. J., Furuta, G. T., Levy, J., VandeWater, J., Whitaker, A. H., Atkins, D., Bauman, M. L., Beaudet, A. L., Carr, E. G., Gershon, M. D., Hyman, S. L., Jirapinyo, P., Jyonouchi, H., Kooros, K., Kushak, R., Levitt, P., Levy, S. E.,. . . Winter, H. (2010). Evaluation, diagnosis, and treatment of gastrointestinal disorders in individuals with ASDs: A consensus report. *Pediatrics, 125*, 1-18. http://doi.org/10.1542/peds.2009-1878c

Chang, Y., & Locke, J. (2016). A systematic review of peer-mediated interventions for children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 27*, 1-10. http://doi.org/10.1016/j.rasd.2016.03.010

Charman, T., & Gotham, K. (2013). Measurement issues: Screening and diagnostic instruments for autism spectrum disorders - lessons from research and practise. *Child and Adolescent Mental Health, 18*(1), 52-63. http://doi.org/10.1111/j.1475-3588.2012.00664.x

Charman, T., Baird, G., Simonoff, E., Loucas, T., Chandler, S., Meldrum, D., & Pickles, A. (2007). Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *British Journal of Psychiatry, 191*(6), 554-559. http://doi.org/10.1192/bjp.bp.107.040196

Chronis, A. M., Fabiano, G. A., Gnagy, E. M., Onyango, A. N., Pelham, W. E., Lopez-Williams, A., Chacko, A., Wymbs, B. T., Coles, E. K., & Seymour, K. E. (2004). An evaluation of the summer treatment program for children with attention-deficit/hyperactivity disorder using a treatment withdrawal design. *Behavior Therapy, 35*(3), 561-585. http://doi.org/10.1016/S0005-7894(04)80032-7

Chronis, A. M., Jones, H. A., & Raggi, V. L. (2006). Evidence-based psychosocial treatments for children and adolescents with attention-deficit/hyperactivity disorder. *Clinical Psychology Review, 26*(4), 486-502. http://doi.org/10.1016/j.cpr.2006.01.002

Coker, T. R., Elliott, M. N., Toomey, S. L., Schwebel, D. C., Cuccaro, P., Tortolero Emery, S., Davies, S. L., Visser, S. N., & Schuster, M. A. (2016). Racial and ethnic disparities in ADHD diagnosis and treatment. *Pediatrics, 138*(3), 1-9. http://doi.org/10.1542/peds.2016-0407

Conners, C. K. (2008). *Conners, 3rd ed.* Western Psychological Services.

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale, 2nd ed.* Western Psychological Services.

Cooper, M., Martin, J., Langley, K., Hamshere, M., & Thapar, A. (2014). Autistic traits in children with ADHD index clinical and cognitive problems. *European Child & Adolescent Psychiatry, 23*(1), 23-34. http://doi.org/10.1007/s00787-013-0398-6

Cordier, R., Vilaysack, B., Doma, K., Wilkes-Gillan, S., & Speyer, R. (2018). Peer inclusion in interventions for children with ADHD: A systematic review and meta-analysis. *BioMed Research International, 2018*, 11-21. http://doi.org/10.1155/2018/7693479

Craig, F., Lamanna, A. L., Margari, F., Matera, E., Simone, M., & Margari, L. (2015). Overlap between autism spectrum disorders and attention deficit hyperactivity disorder: Searching for distinctive/common clinical features. *Autism Research, 8*(3), 328-337. http://doi.org/10.1002/aur.1449

Craig, F., Margari, F., Legrottaglie, A., Palumbi, R., de Giambattista, C., & Margari, L. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric Disease and Treatment, 12*, 1191-1202. http://doi.org/10.2147/NDT.S104620

Danielson, M. L., Bitsko, R. H., Ghandour, R. M., Holbrook, J. R., Kogan, M. D., & Blumberg, S. J. (2018). Prevalence of parent-reported ADHD diagnosis and associated treatment among U.S. children and adolescents, 2016. *Journal of Clinical Child and Adolescent Psychology, 47*(2), 199-212. http://doi.org/10.1080/15374416.2017.1417860

Davis, N., & Kollins, S. (2012). Treatment for co-occurring attention deficit/hyperactivity disorder and autism spectrum disorder. *Journal of the American Society for Experimental Neurotherapeutics, 9*(3), 518-530. http://doi.org/10.1007/s13311-012-0126-9

Dawson, G., & Rogers, S. J. (2010). In Dawson G. (Ed.), *Early start Denver model for young children with autism: Promoting language, learning, and engagement*. Guilford Press.

Dellapiazza, F., Audras-Torrent, L., Michelon, C., & Baghdadli, A. (2021). Clinical characteristics of children with ASD and comorbid ADHD: Association with social impairment and externalizing and internalizing behaviours. *Research in Developmental Disabilities, 113*, 1-10. https://doi.org/10.1016/j.ridd.2021.103930

Demopoulos, C., Hopkins, J., & Davis, A. (2013). A comparison of social cognitive profiles in children with autism spectrum disorders and attention-deficit/hyperactivity disorder: A matter of quantitative but not qualitative difference? *Journal of Autism and Developmental Disorders, 43*(5), 1157-1170. http://doi.org/10.1007/s10803-012-1657-y

Devine, S. L., Rapp, J. T., Testa, J. R., Henrickson, M. L., & Schnerch, G. (2011). Detecting changes in simulated events using Partial-Interval recording and momentary time sampling III: Evaluating sensitivity as a function of session length. *Behavioral Interventions, 26*(2), 103-124. http://doi.org/10.1002/bin.328

Dickson, K. S., Suhrheinrich, J., Rieth, S. R., & Stahmer, A. C. (2018). Parent and teacher concordance of child outcomes for youth with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 48*(5), 1423-1435. https://doi.org/10.1007/s10803-017-3382-z

Dunne, Z. (2020). *Behavioral change in a summer treatment program for children with ASD and ADHD* [Unpublished doctoral dissertation]. University of Washington.

DuPaul, G., Power, T. J., Anastopoulos A. D., & Reid, R. (1998). *ADHD rating scale-IV: Checklists, norms, and clinical interpretation*. Guilford Press.

Duvekot, J., van der Ende, J., Verhulst, F. C., & Greaves-Lord, K. (2014). The screening accuracy of the parent and teacher-reported social responsiveness scale (SRS):

Comparison with the 3Di and ADOS. *Journal of Autism and Developmental Disorders, 45*(6), 1658-1672. http://doi.org/10.1007/s10803-014-2323-3

Evans, S. W., Owens, J. S., & Bunford, N. (2014). Evidence-based psychosocial treatments for children and adolescents with attention-deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology, 43*(4), 527-551. http://doi.org/10.1080/15374416.2013.850700

Evans, S. W., Langberg, J. M., Egan, T., & Molitor, S. J. (2014). Middle school–based and high school–based interventions for adolescents with ADHD. *Child and Adolescent Psychiatric Clinics of North America, 23*(4), 699-715. https://doi.org/10.1016/j.chc.2014.05.004

Fabiano, G. A., Pelham, W. E., Coles, E. K., Gnagy, E. M., Chronis-Tuscano, A., & O'Connor, B. C. (2009). A meta-analysis of behavioral treatments for attention-deficit/hyperactivity disorder. *Clinical Psychology Review, 29*(2), 129-140. http://doi.org/10.1016/j.cpr.2008.11.001

Fabiano, G. A., Schatz, N. K., & Pelham, W. E. (2014). Summer treatment programs for youth with ADHD. *Child and Adolescent Psychiatric Clinics of North America, 23*(4), 757-773. http://doi.org/10.1016/j.chc.2014.05.012

Factor, R. S., Ryan, S. M., Farley, J. P., Ollendick, T. H., & Scarpa, A. (2017). Does the presence of anxiety and ADHD symptoms add to social impairment in children with autism spectrum disorder? *Journal of Autism and Developmental Disorders, 47*(4), 1122-1134. http://doi.org/10.1007/s10803-016-3025-9

Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: A systematic literature review. *European Child & Adolescent Psychiatry, 22*(6), 329-340. http://doi.org/10.1007/s00787-013-0375-0

Fisher, W. A., Groff, R. A., & Roane, H. S. (2011). Applied behavior analysis: History, philosophy, principles, and basic methods. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 3-16). Guilford Press.

Flouri, E., Midouhas, E., Charman, T., & Sarmadi, Z. (2015). Poverty and the growth of emotional and conduct problems in children with autism with and without comorbid ADHD. *Journal of Autism and Developmental Disorders, 45*(9), 2928-2938. http://doi.org/10.1007/s10803-015-2456-z

Gardenier, N. C., MacDonald, R., & Green, G. (2004). Comparison of direct observational methods for measuring stereotypic behavior in children with autism spectrum disorders. *Research in Developmental Disabilities, 25*(2), 99-118. http://doi.org/10.1016/j.ridd.2003.05.004

Gerber, W., Gerber-von Müller, G., Andrasik, F., Niederberger, U., Siniatchkin, M., Kowalski, J. T., Petermann, U., & Petermann, F. (2012). The impact of a multimodal summer camp training on neuropsychological functioning in children and adolescents with ADHD: An exploratory study. *Child Neuropsychology, 18*(3), 242-255. http://doi.org/10.1080/09297049.2011.599115

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2015). *Behavior rating inventory of executive function, 2nd ed.* Western Psychological Services.

Goldin, R. L., Matson, J. L., Tureck, K., Cervantes, P. E., & Jang, J. (2013). A comparison of tantrum behavior profiles in children with ASD, ADHD and comorbid ASD and

ADHD. *Research in Developmental Disabilities, 34*(9), 2669-2675. http://doi.org/10.1016/j.ridd.2013.04.022

Goldstein, A. P., McGinnis, E., Sprafkin, R. P., Gershaw, N. J., & Klein, P. (1997). *Skillstreaming the adolescent: New strategies and perspectives for teaching prosocial skills* (Rev. ed.). Research Press.

Gooch, D., Maydew, H., Sears, C., & Norbury, C. F. (2017). Does a child's language ability affect the correspondence between parent and teacher ratings of ADHD symptoms? *BMC Psychiatry, 17*(1), 1-10. http://doi.org/10.1186/s12888-017-1300-8

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*(5), 581-586. http://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Green, I. W., Kidd, C. L., & Accordino, R. E. (2016). Differential diagnosis of autism spectrum disorder across the lifespan. In C. J. McDougle (Ed.), *Autism spectrum disorder* (pp. 47-66). Oxford University Press.

Groom, M. J., Kochhar, P., Hamilton, A., Liddle, E. B., Simeou, M., & Hollis, C. (2017). Atypical processing of gaze cues and faces explains comorbidity between autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD). *Journal of Autism and Developmental Disorders, 47*(5), 1496-1509. http://doi.org/10.1007/s10803-017-3078-4

Grzadzinski, R., Di Martino, A., Brady, E., Mairena, M., O'Neale, M., Petkova, E., Lord, C., & Castellanos, F. (2011). Examining autistic traits in children with ADHD: Does the autism spectrum extend to ADHD? *Journal of Autism and Developmental Disorders, 41*(9), 1178-1191. http://doi.org/10.1007/s10803-010-1135-3

Grzadzinski, R., Dick, C., Lord, C., & Bishop, S. (2016). Parent-reported and clinician-observed autism spectrum disorder (ASD) symptoms in children with attention deficit/hyperactivity disorder (ADHD): Implications for practice under DSM-5. *Molecular Autism, 7*(4), 1-12. http://doi.org/10.1186/s13229-016-0072-1

Handen, B. L., Aman, M. G., Arnold, L. E., Hyman, S. L., Tumuluru, R. V., Lecavalier, L., Corbett-Dick, P., Pan, X., Hollway, J. A., Buchan-Page, K. A., Silverman, L. B., Brown, N. V., Rice, R. R., Hellings, J., Mruzek, D. W., McAuliffe-Bellin, S., Hurt, E. A., Ryan, M. M., Levato, L., & Smith, T. (2015). Atomoxetine, parent training, and their combination in children with autism spectrum disorder and attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 54*(11), 905-915. http://doi.org/10.1016/j.jaac.2015.08.013

Harrison, J. R., Soares, D. A., Rudzinski, S., & Johnson, R. (2019). Attention deficit hyperactivity disorders and classroom-based interventions: Evidence-based status, effectiveness, and moderators of effects in single-case design research. *Review of Educational Research, 89*(4), 569-611. http://doi.org/10.3102/0034654319857038

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165-179. http://doi.org/ 10.1177/001440290507100203

Hoza, B., Kaiser, N., & Hurt, E. (2007). Multimodal treatments for childhood attention-deficit/hyperactivity disorder: Interpreting outcomes in the context of study designs. *Clinical Child and Family Psychology Review, 10*(4), 318-334. http://doi.org/10.1007/s10567-007-0025-5

Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry, 54*(2), 216-224. http://doi.org/10.1111/j.1469-7610.2012.02589.x

Hyman, S. L., Levy, S. E., Myers, S. M., & AAP Council on Children with Disabilities, Section on Developmental and Behavioral Pediatrics. (2020). Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics, 145*(1), 1-71. http://doi.org/ 10.1542/peds.2019-3447

Iadarola, S., & Smith, T. (2016). Behavioral treatment of autism spectrum disorder. In C. J. McDougle (Ed.), *Autism spectrum disorder* (pp. 253-274). Oxford University Press.

Jang, J., Matson, J. L., Williams, L. W., Tureck, K., Goldin, R. L., & Cervantes, P. E. (2013). Rates of comorbid symptoms in children with ASD, ADHD, and comorbid ASD and ADHD. *Research in Developmental Disabilities, 34*(8), 2369-2378. http://doi.org/10.1016/j.ridd.2013.04.021

Joshi, G., Faraone, S. V., Wozniak, J., Tarko, L., Fried, R., Galdo, M., Furtak, S. L., & Biederman, J. (2017). Symptom profile of ADHD in youth with high-functioning autism spectrum disorder: A comparative study in psychiatrically referred populations. *Journal of Attention Disorders, 21*(10), 846-855. http://doi.org/10.1177/1087054714543368

Joshi, G., Petty, C. R., Fried, R., Wozniak, J., Micco, J. A., Henin, A., Doyle, R., Galdo, M., Kotarski, M., Caruso, J., Meller, B., Faraone, S. V., & Biederman, J. (2011). Discriminant and concurrent validity of a simplified DSM-based structured diagnostic instrument for the assessment of autism spectrum disorders in youth and young adults. *BMC Psychiatry, 11*(1), 1-7. http://doi.org/10.1186/1471-244X-11-204

Kasari, C., & Smith, T. (2016). Forest for the trees: Evidence-Based practices in ASD. *Clinical Psychology: Science and Practice, 23*(3), 260-264. http://doi.org/10.1111/cpsp.12161

Kelley, M. E., LaRue, R. H., Roane, H. S., & Gadaire, D. M. (2011). Indirect behavioral assessments: Interview and rating scales. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 182-190). Guilford Press.

Kern, J. K., Geier, D. A., Sykes, L. K., Geier, M. R., & Deth, R. C. (2012). Are ASD and ADHD a continuum? A comparison of pathophysiological similarities between the disorders. *Journal of Attention Disorders, 19*(9), 805-827. http://doi.org/10.1177/1087054712459886

Kernder, T., Doepfner, M., Dose, C., & Goertz-Dorten, A. (2019). Psychometric properties of a modified version of the Weiss functional impairment rating scale-parent report (WFIRS-P) in a clinical sample of children with aggressive behavior. *Quality of Life Research, 28*(1), 241-251. http://doi.org/10.1007/s11136-018-2015-0

Kiani, B., Hadianfard, H., & Weiss, M. D. (2018). Descriptive and psychometric properties of the Persian version of the Weiss functional impairment rating scale: Parent report form in Iranian children. *Health and Quality of Life Outcomes, 16*(1), 1-9. http://doi.org/10.1186/s12955-018-1053-1

Konold, T. R., Walthall, J. C., & Pianta, R. C. (2019). The behavior of child behavior ratings: Measurement structure of the child behavior checklist across time, informants, and child gender. *Behavioral Disorders, 29*(4), 372-383. https://doi.org/10.1177/019874290402900405

Kretzmann, M., Shih, W., & Kasari, C. (2014). Improving peer engagement of children with autism on the school playground: A randomized controlled trial. *Behavior Therapy, 46*(1), 20-28. http://doi.org/10.1016/j.beth.2014.03.006

Lane, J. D., & Ledford, J. R. (2014). Using interval-based systems to measure behavior in early childhood special education and early intervention. *Topics in Early Childhood Special Education, 34*(2), 83-93. doi:10.1177/0271121414524063

LeBlanc, L. A., Lund, C., Kooken, C., Lund, J. B., & Fisher, W. W. (2020). Procedures and accuracy of discontinuous measurement of problem behavior in common practice of applied behavior analysis. *Behavior Analysis in Practice, 13*(2), 411-420. http://doi.org/10.1007/s40617-019-00361-6

Leitner, Y. (2014). The co-occurrence of autism and attention deficit hyperactivity disorder in children - what do we know? *Frontiers in Human Neuroscience, 8*(1) http://doi.org/10.3389/fnhum.2014.00268

Levy, A., & Perry, A. (2011). Outcomes in adolescents and adults with autism: A review of the literature. *Research in Autism Spectrum Disorders, 5*(4), 1271-1282. https://doi.org/10.1016/j.rasd.2011.01.023

Lindor, E., Sivaratnam, C., May, T., Stefanac, N., Howells, K., & Rinehart, N. (2019). Problem behavior in autism spectrum disorder: Considering core symptom severity and accompanying sleep disturbance. *Frontiers in Psychiatry, 10*, 1-10. http://doi.org/10.3389/fpsyt.2019.00487

Lopata, C., Thomeer, M. L., Volker, M. A., Nida, R. E., & Lee, G. K. (2007). Effectiveness of a manualized summer social treatment program for high-functioning children with autism

spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(5), 890-904. http://doi.org/10.1007/s10803-007-0460-7

Lopata, C., Thomeer, M., Volker, M., Toomey, J., & Nida, R. (2006). Effectiveness of a cognitive-behavioral treatment on the social behaviors of children with Asperger disorder. *Focus on Autism and Other Developmental Disabilities, 21*(4), 237-244. https://doi.org/10.1177/10883576060210040501

Lopata, C., Thomeer, M., Volker, M., Toomey, J., Nida, R., Lee, G., Smerbeck, A. M., & Rodgers, J. (2010). RCT of a manualized social treatment for high-functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders, 40*(11), 1297-1310. http://doi.org/10.1007/s10803-010-0989-8

Lopez-Williams, A., Chacko, A., Wymbs, B. T., Fabiano, G. A., Seymour, K. E., Gnagy, E. M., Chronis, A. M., Burrows-Maclean, L., Pelham, W. E., & Morris, T. L. (2005). Athletic performance and social behavior as predictors of peer acceptance in children diagnosed with attention-deficit/hyperactivity disorder. *Journal of Emotional and Behavioral Disorders, 13*(3), 173-180. https://doi.org/10.1177/10634266050130030501

Lord, C., & Rutter, M. (2012). *Autism diagnostic observation schedule, 2nd ed.* Western Psychological Services.

Lovaas, O. I. (1981). *Teaching developmentally disabled children: The "me" book*. University Park Press.

Maenner, M. J., Shaw, K. A., Baio, J., Washington, A., Patrick, M., DiRenzio, M., Christensen, D. L., Wiggins, L. D., Pettygrove, S., Andrews, J. G., Lopez, M., Hudson, A., Baroud, T., Schwenk, Y., White, T., Rosenberg, C. R., Lee, L., Harrington, R. A., Huston, M.,… Dietz, P. M. (2020). *Prevalence of autism spectrum disorder among children aged 8 years*

– autism and developmental disabilities monitoring network, 11 sites, United States, 2016 (Surveillance Summaries 4, Issue 69). Centers for Disease Control and Prevention. http://doi.org/10.15585/mmwr.ss6904a1

Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A., & Skuse, D. (2011). Sex differences in autism spectrum disorder: Evidence from a large sample of children and adolescents. *Journal of Autism and Developmental Disorders, 42*(7), 1304-1313. http://doi.org/10.1007/s10803-011-1356-0

Martinez, J. R., Waters, C. L., Conroy, M. A., & Reichow, B. (2019). Peer-mediated interventions to address social competence needs of young children with ASD: Systematic review of single-case research design studies. *Topics in Early Childhood Special Education*, 1-13. http://doi.org/10.1177/0271121419839136

Masi, G., Pisano, S., Milone, A., & Muratori, P. (2015). Child behavior checklist dysregulation profile in children with disruptive behavior disorders: A longitudinal study. *Journal of Affective Disorders, 186*, 249-253. http://doi.org/10.1016/j.jad.2015.05.069

May, T., Brignell, A., Hawi, Z., Brereton, A., Tonge, B., Bellgrove, M., & Rinehart, N. (2018). Trends in the overlap of autism spectrum disorder and attention deficit hyperactivity disorder: Prevalence, clinical management, language and genetics. *Current Developmental Disorders Reports, 5*(1), 49-57. http://doi.org/10.1007/s40474-018-0131-8

Mayes, S. D. (2012). *Checklist for autism spectrum disorder*. Stoelting.

Mayes, S. D., Calhoun, S. L., Mayes, R. D., & Molitoris, S. (2012). Autism and ADHD: Overlapping and discriminating symptoms. *Research in Autism Spectrum Disorders, 6*(1), 277-285. http://doi.org/10.1016/j.rasd.2011.05.009

McGrew, J. H., Ruble, L. A., & Smith, I. M. (2016). Autism spectrum disorder and evidence-based practice in psychology. *Clinical Psychology: Science and Practice, 23*(3), 239-255. http://doi.org/10.1111/cpsp.12160

McKeown, C., Hisle-Gorman, E., Eide, M., Gorman, G. H., & Nylund, C. M. (2013). Association of constipation and fecal incontinence with attention-deficit/hyperactivity disorder. *Pediatrics, 132*(5), e1210-e1215. http://doi.org/10.1542/peds.2013-1580

McLeod, B. D., Wood, J. J., & Klebanoff, S. (2015). Advances in evidence-based intervention and assessment practices for youth with an autism spectrum disorder. *Behavior Therapy, 46*(1), 1-6. http://doi.org/10.1016/j.beth.2014.07.004

Mikami, A. Y., Calhoun, C. D., & Abikoff, H. B. (2010). Positive illusory bias and response to behavioral treatment among children with attention-deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology, 39*, 373-385. doi:10.1080/15374411003691735

Mikami, A., Smit, S., & Khalis, A. (2017). Social skills training and ADHD—What works? *Current Psychiatry Reports, 19*(12), 1-9. http://doi.org/10.1007/s11920-017-0850-2

Miodovnik, A., Harstad, E., Sideridis, G., & Huntington, N. (2015). Timing of the diagnosis of attention-deficit/hyperactivity disorder and autism spectrum disorder. *Pediatrics, 136*(4), e830-e837. http://doi.org/10.1542/peds.2015-1502

Mitchell, E., Mrug, S., Patterson, C., Bailey, K., & Hodgens, J. B. (2015). Summer treatment program improves behavior of children with high-functioning autism spectrum disorder. *Journal of Autism and Developmental Disorders, 45*(8), 2295-2310. http://doi.org/10.1007/s10803-014-2241-4

Mrug, S., & Hodgens, J. B. (2008). Behavioral summer treatment program improves social and behavioral functioning of four children with Asperger's disorder. *Clinical Case Studies, 7*(3), 171-190. http://doi.org/10.1177/1534650107306293

MTA Cooperative Group. (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry, 56*(12), 1073-1086. http://doi.org/10.1001/archpsyc.56.12.1073

MTA Cooperative Group. (2004a). National institute of mental health multimodal treatment study of ADHD follow-up: 24-month outcomes of treatment strategies for attention-deficit/hyperactivity disorder. *Pediatrics, 113*(4), 754-761. http://doi.org/10.1542/peds.113.4.754

MTA Cooperative Group. (2004b). National institute of mental health multimodal treatment study of ADHD follow-up: Changes in effectiveness and growth after the end of treatment. *Pediatrics, 113*(4), 762-769. http://doi.org/10.1542/peds.113.4.762

Murray, M. J., Mayes, S. D., & Smith, L. A. (2011). Brief report: Excellent agreement between two brief autism scales (checklist for autism spectrum disorder and social responsiveness scale) completed independently by parents and the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders, 41*(11), 1586-1590. http://doi.org/10.1007/s10803-011-1178-0

National Institute for Children's Health Quality, & American Academy of Pediatrics. (2002) *NICHQ Vanderbilt Assessment Scales.* https://www.nichq.org/sites/default/files/resource-file/NICHQ_Vanderbilt_Assessment_Scales.pdf

National Institute of Mental Health. (2017, November). Attention deficit/hyperactivity disorder (ADHD). National Institute of Mental Health.

https://www.nimh.nih.gov/health/statistics/attention-deficit-hyperactivity-disorder-adhd.shtml

National Research Council, Committee on Educational Interventions for Children with Autism. (2001). In C. Lord & J. P. McGee (Eds.), *Educating children with autism.* National Academy Press. http://doi.org/10.17226/10017

O'Connor, B., Fabiano, G., Waschbusch, D., Belin, P., Gnagy, E., Pelham, W., Greiner, A. R., & Roemmich, J. (2014). Effects of a summer treatment program on functional sports outcomes in young children with ADHD. *Journal of Abnormal Child Psychology; an Official Publication of the International Society for Research in Child and Adolescent Psychopathology, 42*(6), 1005-1017. http://doi.org/10.1007/s10802-013-9830-0

Odom, S. L. (2019). Peer-based interventions for children and youth with autism spectrum disorder: History and effects. *School Psychology Review, 48*(2), 170-176. http://doi.org/10.17105/SPR-2019-0019.V48-2

Odom, S. L., Collet-Klingenberg, L., Rogers, S. J., & Hatton, D. D. (2010). Evidence-based practices in interventions for children and youth with autism spectrum disorders. *Preventing School Failure: Alternative Education for Children and Youth: Autism Spectrum Disorders, 54*(4), 275-282. http://doi.org/10.1080/10459881003785506

Owens, E., Hinshaw, S., Kraemer, H. C., Arnold, L., Abikoff, H., Cantwell, D., Conners, C. K., Elliott, G., Greenhill, L. L., Hechtman, L., Hoza, B., Jensen, P. S., March, J. S., Newcorn, J. H., Pelham, W. E., Severe, J. B., Swanson, J. M., Vitiello, B., Wells, K. C., & Wigal, T. (2003). Which treatment for whom for ADHD? Moderators of treatment response in the MTA. *Journal of Consulting and Clinical Psychology, 71*(3), 540-552. http://doi.org/10.1037/0022-006X.71.3.540

Pandolfi, V., Magyar, C. I., & Dill, C. A. (2012). An initial psychometric evaluation of the CBCL 6–18 in a sample of youth with autism spectrum disorders. *Research in Autism Spectrum Disorders, 6*(1), 96-108. http://doi.org/10.1016/j.rasd.2011.03.009

Papachristou, E., Schulz, K., Newcorn, J., Bédard, A.,V., Halperin, J. M., & Frangou, S. (2016). Comparative evaluation of child behavior checklist-derived scales in children clinically referred for emotional and behavioral dysregulation. *Frontiers in Psychiatry, 7*, 1-8. http://doi.org/10.3389/fpsyt.2016.00146

Papadopoulos, N., Sciberras, E., Hiscock, H., Mulraney, M., Mcgillivray, J., & Rinehart, N. (2019). The efficacy of a brief behavioral sleep intervention in school-aged children with ADHD and comorbid autism spectrum disorder. *Journal of Attention Disorders, 23*(4), 341-350. http://doi.org/10.1177/1087054714568565

Pecora, P. J., Amsden, G., Le Prohn, N. S., & Achenbach, T. M. (2002). Conclusions and recommendations for future research. In N. S. Le Prohn, K. M., Wetherbee, E. R. Lamont, T. M. Achenbach, P. J., Pecora (Eds.), *Assessing Youth Behavior Using the Child Behavior Checklist in Family and Children's Services* (pp. 181-194). CWLA Press.

Pelham, W. E., & Fabiano, G. A. (2008). Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology, 37*(1), 184-214. http://doi.org/10.1080/15374410701818681

Pelham, W. E., Fabiano, G. A., & Massetti, G. M. (2005). Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*(3), 449-476. http://doi.org/10.1207/s15374424jccp3403_5

Pelham, W. E., Greiner, A. R., & Gnagy, E. M. (2012). *Children's summer treatment program manual.* Florida International University.

Pelham, W., Gnagy, E., Greiner, A., Hoza, B., Hinshaw, S., Swanson, J., Simpson, S., Shapiro, C., Bukstein, O., Baron-Myak, C., & McBurnett, K. (2000). Behavioral versus behavioral and pharmacological treatment in ADHD children attending a summer treatment program. *Journal of Abnormal Child Psychology; an Official Publication of the International Society for Research in Child and Adolescent Psychopathology, 28*(6), 507-525. http://doi.org/1005127030251

Piper, B. J., Gray, H. M., Raber, J., & Birkett, M. A. (2014). Reliability and validity of brief problem monitor, an abbreviated form of the child behavior checklist. *Psychiatry and Clinical Neurosciences, 68*(10), 759-767. http://doi.org/10.1111/pcn.12188

Radley, K. C., O' Handley, R. D., & Labrot, Z. C. (2015). A comparison of momentary time sampling and partial-interval recording for assessment of effects of social skills training. *Psychology in the Schools, 52*(4), 363-378. http://doi.org/10.1002/pits.21829

Ramirez, L. F., & Sajatovic, M. (2012). In Ramirez L. F. (Ed.), *Rating scales in mental health* (3rd ed.). Johns Hopkins University Press.

Rao, P. A., & Landa, R. J. (2014). Association between severity of behavioral phenotype and comorbid attention deficit hyperactivity disorder symptoms in children with autism spectrum disorders. *Autism, 18*(3), 272-280. http://doi.org/10.1177/1362361312470494

Rao, P. A., Beidel, D. C., & Murray, M. J. (2008). Social skills interventions for children with Asperger's syndrome or high-functioning autism: A review and recommendations. *Journal of Autism and Developmental Disorders, 38*(2), 353–361. http://doi.org/10.1007/s10803-007-0402-4

Rapp, J. T., Carroll, R. A., Stangeland, L., Swanson, G., & Higgins, W. J. (2011). A comparison of reliability measures for continuous and discontinuous recording methods: Inflated agreement scores with partial interval recording and momentary time sampling for duration events. *Behavior Modification, 35*(4), 389-402. http://doi.org/10.1177/0145445511405512

Re, A. M., & Cornoldi, C. (2009). Two new rating scales for assessment of ADHD symptoms in Italian preschool children. *Journal of Attention Disorders, 12*(6), 532-539. http://doi.org/10.1177/1087054708323001

Reed, M., Jakubovski, E., Johnson, J., & Bloch, M. (2017). Predictors of long-term school-based behavioral outcomes in the multimodal treatment study of children with attention-deficit/hyperactivity disorder. *Journal of Child and Adolescent Psychopharmacology, 27*(4), 296-309. http://doi.org/10.1089/cap.2015.0168

Reiersen, A. M., & Todd, R. D. (2008). Co-occurrence of ADHD and autism spectrum disorders: Phenomenology and treatment. *Expert Review of Neurotherapeutics, 8*(4), 657-669. http://doi.org/10.1586/14737175.8.4.657

Reiersen, A. M., Constantino, J. N., Volk, H. E., & Todd, R. D. (2007). Autistic traits in a population-based ADHD twin sample. *Journal of Child Psychology and Psychiatry, 48*(5), 464-472. http://doi.org/10.1111/j.1469-7610.2006.01720.x

Repp, A. C., Roberts, D. M., Slack, D. J., Repp, C. F., & Berkler, M. S. (1976). A comparison of frequency, interval, and time-sampling methods of data collection. *Journal of Applied Behavior Analysis, 9*(4), 501-508. http://doi.org/10.1901/jaba.1976.9-501

Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior assessment system for children, 3rd ed.* Pearson.

Roane, H. S., Rihgdahl, J. E., Kelley, M. E., & Glover, A. C. (2011). Single-case experimental designs. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 132-150). Guilford Press.

Robins, D., Fein, D., & Barton, M. (2009). *Modified checklist for autism in toddlers, revised with follow-up.* https://m-chat.org/en-us/page/take-m-chat-test/online

Rommelse, N., Visser, J., & Hartman, C. (2018). Differentiating between ADHD and ASD in childhood: Some directions for practitioners. *European Child & Adolescent Psychiatry, 27*(6), 679-681. http://doi.org/10.1007/s00787-018-1165-5

Ros, R., & Graziano, P. A. (2020). A transdiagnostic examination of self-regulation: comparisons across preschoolers with ASD, ADHD, and typically developing children. *Journal of Clinical Child and Adolescent Psychology, 49*(4), 493-508. https://doi.org/10.1080/15374416.2019.1591280

Ros-Demarize, R., & Graziano, P. A. (2021). Initial feasibility and efficacy of the summer treatment program (STP-PreK) for preschoolers with autism spectrum disorder and comorbid externalizing behavior problems. *Journal of Early Intervention, 43*(1), 60-79. https://doi.org/10.1177/1053815120917452

Rucklidge, J. J. (2010). Gender differences in attention-deficit/hyperactivity disorder. *Psychiatric Clinics of North America, 33*(2), 357-373. http://doi.org/10.1016/j.psc.2010.01.006

Rutter, M., Bailey, A., & Lord, C. (2003). *Social communication questionnaire.* Western Psychological Services.

Rutter, M., LeCouteur A., & Lord, C. (2003). *Autism diagnostic interview - revised*. Western Psychological Services.

Sam, A., Reszka, S., Odom, S., Hume, K., & Boyd, B. (2015). The use of coding methods to estimate the social behavior directed toward peers and adults of preschoolers with ASD in TEACCH, LEAP, and eclectic "BAU" classrooms. *Behavioral Disorders, 40*(2), 91-101. http://doi.org/10.17988/BD-13-47.1

Santosh, P. J., Baird, G., Pityaratstian, N., Tavare, E., & Gringras, P. (2006). Impact of comorbid autism spectrum disorders on stimulant response in children with attention deficit hyperactivity disorder: A retrospective and prospective effectiveness study. *Child: Care, Health and Development, 32*(5), 575-583. http://doi.org/10.1111/j.1365-2214.2006.00631.x

Sasser, T. R., Kalvin, C. B., & Bierman, K. L. (2016). Developmental trajectories of clinically significant attention-deficit/hyperactivity disorder (ADHD) symptoms from grade 3 through 12 in a high-risk sample: Predictors and outcomes. *Journal of Abnormal Psychology, 125*(2), 207-219. http://doi.org/10.1037/abn0000112

Sayal, K., & Taylor, E. (2005). Parent ratings of school behaviour in children at risk of attention deficit/hyperactivity disorder. *Acta Psychiatrica Scandinavica, 111*(6), 460-465. http://doi.org/10.1111/j.1600-0447.2004.00487.x

Schoenfelder, E. N., & Sasser, T. (2016). Skills versus pills: Psychosocial treatments for ADHD in childhood and adolescence. *Pediatric Annals, 45*(10), 367-372. http://doi.org/10.3928/19382359-20160920-04

Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *Childhood autism rating scale, 2nd ed.* Western Psychological Services.

Scott, N., Blair, P. S., Emond, A. M., Fleming, P. J., Humphreys, J. S., Henderson, J., & Gringras, P. (2013). Sleep patterns in children with ADHD: A population-based cohort study from

birth to 11 years. *Journal of Sleep Research, 22*(2), 121-128. http://doi.org/10.1111/j.1365-2869.2012.01054.x

Sellinger, V. J., & Elder, J. H. (2016). Parent training intervention to manage externalizing behaviors in children with autism. *Journal of Pediatric Health Care, 30*(1), 73-77. http://doi.org/10.1016/j.pedhc.2015.10.002

Shadish, W., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage Learning.

Sidol, C., & Epstein, J. (2020). Identifying potential moderators of ADHD summer treatment program outcomes: A systematic review. *The ADHD Report, 28*(3), 1-9. https://doi.org/10.1521/adhd.2020.28.3.1

Sims, D. M., & Lonigan, C. J. (2012). Multi-method assessment of ADHD characteristics in preschool children: Relations between measures. *Early Childhood Research Quarterly, 27*(2), 329-337. http://doi.org/10.1016/j.ecresq.2011.08.004

Skuse, D., Warrington, R., Bishop, D., Chowdhury, U., Lau, J., & Mandy, W. (2004). The developmental, dimensional and diagnostic interview (3di): A novel computerized assessment for autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(5), 548–558. https://doi.org/10.1097/00004583-200405000-00008

Smith, T., Aman, M. G., Arnold, L. E., Silverman, L. B., Lecavalier, L., Hollway, J., Tumuluru, R., Hyman, S. L., Buchan-Page, K. A., Hellings, J., Rice, R. R., Brown, N. V., Pan, X., & Handen, B. L. (2016). Atomoxetine and parent training for children with autism and attention-deficit/hyperactivity disorder: A 24-week extension study. *Journal of the*

American Academy of Child & Adolescent Psychiatry, 55(10), 868-876.e2. http://doi.org/10.1016/j.jaac.2016.06.015

Stokes, J., Pogge, D., Wecksell, B., & Zaccario, M. (2011). Parent–child discrepancies in report of psychopathology: The contributions of response bias and parenting stress. *Journal of Personality Assessment, 93*(5), 527-536. http://doi.org/10.1080/00223891.2011.594131

Strain, P. S., & Bovey, E. H. (2011). Randomized, controlled trial of the LEAP model of early intervention for young children with autism spectrum disorders. *Topics in Early Childhood Special Education, 31*(3), 133-154. http://doi.org/10.1177/0271121411408740

SungWoo, K., Ingvarsson, E. T., Quigg, A. M., Seckinger, K. E., & Teichman, H. M. (2011). Defining and measuring behavior. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 3-16). Guilford Press.

Tabachnick, B., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Pearson Education.

Tarakçıoğlu, M., Çakın Memik, N., Olgun, N., Aydemir, Ö., & Weiss, M. (2015). Turkish validity and reliability study of the Weiss functional impairment rating scale-parent report. *ADHD Attention Deficit and Hyperactivity Disorders, 7*(2), 129-139. http://doi.org/10.1007/s12402-014-0158-6

Tatsuoka, C. (2012). The statistical evaluation of outcome data. In M. Sajatovic & L. F. Ramirez (Eds.), *Rating scales in mental health* (3rd ed., pp. 11-14). Johns Hopkins University Press.

Taylor, M. J., Charman, T., & Ronald, A. (2015). Where are the strongest associations between autistic traits and traits of ADHD? evidence from a community-based twin study. *European Child & Adolescent Psychiatry, 24*(9), 1129-1138. http://doi.org/10.1007/s00787-014-0666-0

Thompson, R. H., & Borrero, J. C. (2011). Direct observation. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 3-16). Guilford Press.

Tripp, G., Schaughency, E., & Clarke, B. (2006). Parent and teacher rating scales in the evaluation of attention-deficit hyperactivity disorder: Contribution to diagnosis and differential diagnosis in clinically referred children. *Journal of Developmental and Behavioral Pediatrics, 27*(3), 209-219. http://doi.org/10.1097/00004703-200606000-00006

Valo, S., & Tannock, R. (2010). Diagnostic instability of DSM–IV ADHD subtypes: Effects of informant source, instrumentation, and methods for combining symptom reports. *Journal of Clinical Child and Adolescent Psychology, 39*(6), 749-760. http://doi.org/10.1080/15374416.2010.517172

van der Meer, Jolanda M. J, Oerlemans, A. M., van Steijn, D.,J., Lappenschaar, M. G. A., de Sonneville, Leo M. J, Buitelaar, J. K., & Rommelse, N. N. J. (2012). Are autism spectrum disorder and attention-deficit/hyperactivity disorder different manifestations of one overarching disorder? Cognitive and symptom evidence from a clinical and population-based sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(11), 1160-1172. http://doi.org/10.1016/j.jaac.2012.08.024

van Wijngaarden-Cremers, P. J.,M., van Eeten, E., Groen, W. B., Van Deurzen, P.,A., Oosterling, I. J., & Van der Gaag, R. J. (2013). 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*(3), 627-635. http://doi.org/10.1007/s10803-013-1913-9

Volkmar, F., Siegel, M., Woodbury-Smith, M., King, B., McCracken, J., State, M., & American Academy of Child and Adolescent Psychiatry Committee on Quality Issues. (2014). Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 53*(2), 237-257. http://doi.org/10.1016/j.jaac.2013.10.013

Volpe, R. J., & Gadow, K. D. (2010). Creating abbreviated rating scales to monitor classroom inattention-overactivity, aggression, and peer conflict: Reliability, validity, and treatment sensitivity. *School Psychology Review, 39*(3), 350-363. http://doi.org/10.1080/02796015.2010.12087757

Volpe, R. J., McConaughy, S. H., & Hintze, J. M. (2009). Generalizability of classroom behavior problem and on-task scores from the direct observation form. *School Psychology Review, 38*(3), 382-401. http://doi.org/10.1080/02796015.2009.12087822

Wells, K., Pelham, W., Kotkin, R., Hoza, B., Abikoff, H., Abramowitz, A., Arnold, L. E., Cantwell, D. P., Conners, C. K., Del Carmen, R., Elliott, G., Greenhill, L. L., Hechtman, L., Hibbs, E., Hinshaw, S. P., Jensen, P. S., March, J. S., Swanson, J. M., & Schiller, E. (2000). Psychosocial treatment strategies in the MTA study: Rationale, methods, and critical issues in design and implementation. *Journal of Abnormal Child Psychology, 28*(6), 483-505. http://doi.org/10.1023/A:1005174913412

Wetherbee, K. M., & Achenbach, T. M. (2002). Research with the CBCL: Methodological and statistical issues. In N. S. Le Prohn, K. M., Wetherbee, E. R. Lamont, T. M. Achenbach, P. J., Pecora (Eds.), *Assessing youth behavior using the child behavior checklist in family and children's services* (pp. 171-179). CWLA Press.

Wigham, S., McConachie, H., Tandos, J., & Le Couteur, A. S. (2012). The reliability and validity of the social responsiveness scale in a UK general child population. *Research in Developmental Disabilities: A Multidisciplinary Journal, 33*(3), 944-950. http://doi.org/10.1016/j.ridd.2011.12.017

Williams White, S., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of Autism and Developmental Disorders, 37*(10), 1858-1868. http://doi.org/10.1007/s10803-006-0320-x

Wilson, B. J., Berg, J. L., Zurawski, M. E., & King, K. A. (2013). Autism and externalizing behaviors: Buffering effects of parental emotion coaching. *Research in Autism Spectrum Disorders, 7*(6), 767-776. http://doi.org/10.1016/j.rasd.2013.02.005

Wilson, B. J., Manangan, C. N., Dauterman, H. A., & Davis, H. N. (2014). ADHD symptoms moderate the relation between ASD status and internalizing symptoms in 3–6-year-old children. *Journal of Autism and Developmental Disorders, 44*(6), 1347-1356. http://doi.org/10.1007/s10803-013-1995-4

Wing, L., Leekam, S. R., Libby, S. J., Gould, J., & Larcombe, M. (2002). The diagnostic interview for social and communication disorders: Background, inter-rater reliability and clinical use. *Journal of Child Psychology and Psychiatry, 43*(3), 307-325. http://doi.org/10.1111/1469-7610.00023

Wirth, O., Slaven, J., & Taylor, M. A. (2014). Interval sampling methods and measurement error: A computer simulation. *Journal of Applied Behavior Analysis, 47*(1), 83-100. http://doi.org/10.1002/jaba.93

Wong, C., Odom, S., Hume, K., Cox, A., Fettig, A., Kucharczyk, S., Brock, M. E., Plavnick, J. B., Fleury, V. P., & Schultz, T. (2015). Evidence-based practices for children, youth, and

young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders, 45*(7), 1951-1966. http://doi.org/10.1007/s10803-014-2351-z

Wymbs, B. T., Robb, J. A., Chronis, A. M., Massetti, G. M., Fabiano, G. A., Arnold, F. W., Brice, A., Gnagy, E. M., Pelham, W. E., Burrows-MacLean, L., & Hoffman, M. T. (2005). Long-term, multimodal treatment of a child with Asperger's syndrome and comorbid disruptive behavior problems: A case illustration. *Cognitive and Behavioral Practice; Cognitive and Behavioral Practice, 12*(3), 338-350. http://doi.org/10.1016/S1077-7229(05)80056-2

Xu, G., Strathearn, L., Liu, B., & Bao, W. (2018). Prevalence of autism spectrum disorder among US children and adolescents, 2014-2016. *Journal of the American Medical Association, 319*, 81-82. http://doi.org/10.1001/jama.2017.17812

Yerys, B. E., Nissley-Tsiopinis, J., de Marchena, A., Watkins, M. W., Antezana, L., Power, T. J., & Schultz, R. T. (2017). Evaluation of the ADHD rating scale in youth with autism. *Journal of Autism and Developmental Disorders, 47*(1), 90-100. http://doi.org/10.1007/s10803-016-2933-z

Table 1

Demographic Characteristics of Apex Participants

Characteristic	*n*	Percentage
Gender		
Male	95	81.2%
Female	22	18.8%
Ethnicity		
White	77	65.8%
Multiracial	15	12.8%
Asian	12	10.3%
Other	12	10.3%
Returner	50	42.7%
Diagnosis		
ADHD	50	42.7%
ASD	39	33.3%
Comorbid	28	23.9%

Note. $N = 117$. Participants average age of 8.12 years ($SD = 1.47$).

Table 2

Behavioral Categories for Coding Defined by the STP Manual

Behavior Name	Definition
Positive Interval Categories	
Following Activity Rules	The child, for the duration of an interval, has not lost any points for Violating Activity Rules.
Good Sportsmanship	The child, for the duration of an interval, (1) has not lost any points for Poor Sportsmanship and (2) has not been in time out for any part of the interval.
Behavior Bonus	The child, for the duration of an interval, (1) has not lost points for any negative point system behaviors with the exception of Violating Activity Rules and Poor Sportsmanship, and (2) has not been in time out for any part of the interval.
Positive Frequency Categories	
Standardized Attention	The child demonstrates that he or she has been paying attention to the ongoing activity by being able to answer correctly within five seconds a staff member's question about the ongoing activity.
Compliance	The child exhibits within ten seconds or within a time specified by the staff member a behavior that has been specified in a command issued by a staff member, or ceases to exhibit for at least ten seconds a behavior the cessation of which has been specified in a command issued by a staff member.
Helping a Peer	The child voluntarily provides to a peer assistance or aid that (1) is relevant to the ongoing activity; (2) is offered in a manner that does not disrupt the ongoing activity or meet the criteria for any negative behavior category; and (3) is accepted.
Sharing with a Peer	The child voluntarily provides to a peer possessions, privileges, or materials that are personal or temporarily assigned to the child, that (1) are relevant to the ongoing activity; (2) are offered in a manner that does not disrupt the ongoing activity or meet the criteria for any negative behavior category, and (3) are accepted.
Contributing to Group Discussions [a]	The child provides a nonredundant rule or name of the weekly social skill to a group discussion.
Novel Contribution [a]	The child provides a task-related, nonredundant statement to group discussion that is not an activity rule or name of the weekly social skill.

Behavior Name	Definition
Ignoring a Negative Stimulus	The child shows no observable negative response to any verbal or nonverbal behavior, from another child, that would typically elicit annoyance or distress from the recipient.
Negative Interval Categories	
Violating Activity Rules	The child violates rules specific to a particular activity, game, or drill.
Poor Sportsmanship	The child, during a game period or skill drill period, (1) does not actively participate in the game or skill drill; (2) inappropriately plays another person's position or hogs the ball; (3) speaks with excessive pride (brags or boasts) about individual, peer, or team performance; (4) complains about the game situation or about individual, peer, or team performance; (5) cheats or attempts to cheat; (6) refuses to share equipment; (7) uses equipment inappropriately; or (8) loses points for Intentional Aggression, Lying, Verbal Abuse to Staff, Name Calling/Teasing, or Cursing/Swearing.
Negative Frequency Categories	
Intentional Aggression	The child apparently intentionally performs a physical behavior that (1) would typically produce physical injury or pain to another, or (2) intrudes on another by inappropriately restricting freedom of movement.
Unintentional Aggression	The child apparently unintentionally performs a physical behavior that (1) would typically produce physical injury or pain to another, or (2) intrudes on another by inappropriately restricting freedom of movement.
Intentional Destruction of Property	The child apparently intentionally performs a physical behavior that (1) destroys an object; (2) damages an object, defaces an object's surface, or otherwise alters an object such that the object's value or usefulness is substantially impaired or substantially reduced at least temporarily; or (3) would typically meet criterion 1 or 2.
Unintentional Destruction of Property	The child apparently unintentionally performs a physical behavior that (1) destroys an object; (2) damages an object, defaces an object's surface, or otherwise alters an object such that the object's value or usefulness is substantially impaired or substantially reduced at least temporarily; or (3) would typically meet criterion 1 or 2.
Noncompliance	The child, when given a command, fails to meet the criteria for Compliance.

Behavior Name	Definition
Repeated Noncompliance	The child, when given a repeated command, fails to meet the criteria for Compliance.
Stealing	The child has possession of an object that belongs to another person without prior permission from the owner of the object or from a staff member.
Leaving the Activity Area	The child leaves the area designated for an activity without permission from a staff member.
Lying	The child reports an event occurring in the program setting that is contradictory to what a staff member knows or suspects to be true.
Verbal Abuse	The child directs a negative communication toward one or more staff members who are identifiable as intended recipients and who can see or hear the negative communication.
Name Calling/Teasing	The child directs a negative communication toward one or more peers who are identifiable as intended recipients and who can see or hear the negative communication.
Cursing/Swearing	The child exhibits any verbal or nonverbal behavior, regardless of tone or intensity, that would typically be regarded as profane, obscene, or offensive and that is not directed toward an individual who can see or hear it.
Interruption	The child exhibits any verbal or nonverbal behavior, with or without meaning, that intrudes into the activity or conversation of others.
Complaining/Whining	The child exhibits any verbal or nonverbal behavior that inappropriately expresses discomfort, dissatisfaction, or resentment, through content, gesture, or tone of voice.

Note. This table summarizes the operational definitions of all behaviors recorded during the STP as presented on pages 17-36 of the STP manual by Pelham et al. (2012).

[a] Contributing to Group Discussion was modified for Apex to only be applicable to contributions that are not original ideas, such as an activity rule or previously identified social skill. The category of Novel Contribution was added to separately record the contribution of new or original ideas. Therefore, definitions provided for these categories are not direct quotations from the manual.

Table 3

Zero-Order Correlations of Apex Child Outcome and Predictor Variables

Measure	*M*	*(SD)*	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.
Outcome																							
1. Contributing Baseline	5.91	(4.84)	--																				
2. Novel Cont. Baseline	1.27	(1.53)	**0.54**	--																			
3. Ignoring Baseline	3.01	(3.71)	**0.31**	**0.33**	--																		
4. IAS Baseline	0.39	(2.57)	-0.10	0.11	0.11	--																	
5. RNC Baseline	0.16	(0.80)	-0.09	0.17	0.17	**0.84**	--																
6. VA Baseline	0.53	(2.09)	-0.10	-0.09	-0.09	**0.26**	0.01	--															
7. Leaving Baseline	0.84	(4.92)	-0.02	**0.27**	**0.27**	0.08	**0.24**	0.11	--														
8. Contributing Final	5.03	(3.66)	**0.39**	0.06	0.06	-0.15	-0.09	-0.06	0.07	--													
9. Novel Cont. Final	4.15	(2.48)	**0.33**	0.06	0.06	-0.04	-0.03	0.12	-0.13	**0.32**	--												
10. Ignoring Final	6.45	(6.82)	**0.25**	**0.46**	**0.46**	0.07	0.13	-0.09	**0.31**	0.07	-0.02	--											
11. IAS Final	0.09	(0.43)	-0.13	0.08	0.08	**0.84**	**0.81**	0.04	**0.21**	-0.11	-0.12	0.03	--										
12. RNC Final	0.06	(0.28)	0.01	**0.21**	**0.21**	-0.04	0.00	-0.04	0.03	-0.01	-0.11	0.11	-0.05	--									
13. VA Final	1.20	(4.84)	-0.10	0.01	0.01	-0.03	-0.02	**0.24**	-0.04	0.14	0.03	0.05	**0.34**	-0.02	--								
14. Leaving Final	0.15	(0.57)	-0.17	-0.01	-0.01	**0.42**	**0.47**	0.12	**0.20**	-0.10	-0.04	-0.01	**0.38**	0.00	0.01	--							
Predictors																							
15. Gender (1 = Male)	0.81	(0.39)	0.10	**0.24**	**0.24**	0.07	0.10	0.13	0.08	0.13	-0.02	**0.19**	0.10	0.03	0.10	0.13	--						
16. Returner (1 = Returner)	0.43	(0.43)	0.01	-0.04	-0.04	-0.02	-0.08	0.09	0.07	-0.09	-0.02	-0.03	-0.01	-0.12	-0.06	0.13	0.11	--					
17. Mean Bunk Age	8.13	(1.37)	0.12	**-0.28**	**-0.28**	-0.03	-0.09	0.17	-0.15	0.02	**0.28**	**-0.23**	-0.05	-0.17	0.01	0.06	-0.15	**0.34**	--				
18. ASD	0.33	(0.47)	0.08	0.00	0.00	-0.11	-0.07	-0.02	-0.03	0.08	-0.03	-0.01	0.03	-0.02	-0.07	-0.09	0.11	0.12	0.08	--			
19. Comorbid	0.43	(0.50)	-0.10	0.00	0.00	0.07	-0.01	0.16	0.14	-0.02	**-0.22**	0.01	-0.02	0.03	0.11	-0.04	0.17	0.04	0.01	**-0.40**	--		
20. SRS-2 Score	64.84	(21.31)	-0.08	-0.12	-0.12	-0.05	-0.09	-0.05	-0.09	0.01	**-0.30**	0.06	0.01	0.13	0.01	-0.09	-0.15	-0.03	0.05	**0.20**	**0.26**	--	
21. CBCL Score	14.60	(8.58)	-0.12	-0.14	-0.14	0.03	0.09	0.10	0.08	0.02	-0.11	0.04	0.07	0.11	0.06	**0.23**	0.04	-0.15	-0.03	**-0.27**	**0.19**	0.18	--

Note. $N = 117$; IAS = Intentional Aggression to Staff, RNC = Repeated Noncompliance, VA = Verbal Abuse, SRS-2 = Social Responsiveness Scale, Second Edition, CBCL = Child Behavior Checklist; categorical variables are dummy coded. First and last timepoints reported, Pearson's r reported. Significant correlations denoted in bold.

Table 4

Means and Standard Deviations of Measures by Child Diagnosis

Measure	ADHD Mean *(SD)* $n = 50$	ASD Mean (*SD)* $n = 39$	Comorbid Mean *(SD)* $n = 28$
Contributing Baseline	5.90 *(4.93)*	6.49 *(5.72)*	5.15 *(4.34)*
Novel Cont. Baseline	1.63 *(1.81)*	0.92 *(1.14)*	1.11 *(1.31)*
Ignoring Baseline	3.00 *(3.84)*	3.03 *(3.74)*	3.00 *(3.57)*
IAS Baseline	0.51 *(3.57)*	0.00 *(0.00)*	0.70 *(2.13)*
RNC Baseline	0.22 *(1.16)*	0.08 *(0.28)*	0.15 *(0.36)*
VA Baseline	0.24 *(1.05)*	0.49 *(1.54)*	1.11 *(3.61)*
Leaving Baseline	0.35 *(1.33)*	0.62 *(3.14)*	2.04 *(9.24)*
Contributing Final	4.80 *(2.91)*	5.42 *(4.23)*	4.89 *(4.10)*
Novel Cont. Final	4.76 *(2.57)*	4.05 *(2.51)*	3.15 *(1.91)*
Ignoring Final	6.41 *(6.72)*	6.39 *(6.74)*	6.59 *(7.35)*
IAS Final	0.08 *(0.45)*	0.11 *(0.51)*	0.07 *(0.27)*
RNC Final	0.06 *(0.32)*	0.05 *(0.23)*	0.07 *(0.27)*
VA Final	1.04 *(2.92)*	0.74 *(2.72)*	2.15 *(8.61)*
Leaving Final	0.22 *(0.80)*	0.08 *(0.27)*	0.11 *(0.32)*
Gender	0.70 *(0.46)*	0.87 *(0.34)*	0.93 *(0.26)*
Returner	0.34 *(0.48)*	0.51 *(0.51)*	0.46 *(0.51)*
Mean Bunk Age	8.00 *(1.34)*	8.28 *(1.37)*	8.17 *(1.43)*
SRS-2	55.32 *(19.82)*	70.71 *(17.90)*	74.92 *(21.67)*
CBCL	15.43 *(7.79)*	11.32 *(6.98)*	17.63 *(10.70)*

Note. N = 117. IAS = Intentional Aggression to Staff, RNC = Repeated Noncompliance, VA = Verbal Abuse, SRS-2 = Social Responsiveness Scale, Second Edition; CBCL = Child Behavior Checklist. Baseline refers to the count of behavior on the first day of camp while final refers to the count of behavior on the last day of camp.

Table 5

Intraclass Correlation Coefficients of Intercept-Only Models

Behavior	Bunk	Child
Contributing to Group Discussion	0.04	0.52
Novel Contribution	0.01	0.39
Ignoring a Negative Stimulus	0.04	0.52
Intentional Aggression - Staff	0.00	0.15
Repeated Noncompliance	0.00	0.07
Verbal Abuse	0.00	0.18
Leaving the Activity Area	0.00	0.09

Note. ICCs of intercept-only models are presented. Bunk served as Level 3 and child served as Level 2. Multiple measurements within children were Level 1.

Table 6

Linear Growth Model Results Without Predictors

	Contributing			Novel Contribution			Ignoring a Negative Stimulus			Intentional Aggression - Staff			Repeated Noncompliance			Verbal Abuse			Leaving the Activity Area		
Fixed Effects	*Coeff*	*SE*	*ES*	*Coeff*	*SE*	*ES*	*Coeff*	*SE*	*ES*	*Coeff*	*SE*	*ES*	*Coeff*	*SE*	*ES*	*Coeff*	*SE*	*ES*	*Coeff*	*SE*	*ES*
Intercept (Baseline)	5.38	0.39 ***	1.28	2.49	0.21 ***	1.10	3.52	1.08 **	0.30	0.42	0.17 *	0.23	0.11	0.03 **	0.34	0.44	0.17 **	0.24	0.55	0.14 ***	0.36
Linear Growth	-0.02	0.01 **	-0.18	0.07	0.01 ***	0.65	0.14	0.01 ***	1.29	0.00	0.01	0.00	0.00	0.00	0.00	0.02	0.01 ***	0.18	-0.01	0.01	-0.09
Random Effects	*Var*			*Var*			*Var*			*Var*			*Var*			*Var*			*Var*		
Bunk Intercept (Baseline)	0.63			0.08			11.05			0.00			0.00			0.00			0.00		
Child Intercept (Baseline)	8.60			3.40			0.03			1.51			0.02			1.68			0.73		
Child Slope (Linear Growth)										0.00			0.00						0.00		
Residual	7.20			4.64			38.71			8.38			0.26			7.76			7.08		
Model Fit																					
Approximate R^2	0.35			0.31			0.20			0.11			0.07			0.13			0.06		
BIC	10711.20			9713.70			13990.90			10848.30			3335.10			10692.20			10440.10		
Deviance (-2LL)	10672.90			9675.30			13952.50			10802.30			3289.10			10653.80			10394.10		
Residual *df*	2136			2136			2137			2136			2136			2137			2136		

Note. $N = 117$. Campers measured at 19 timepoints; time coded in days from baseline (baseline = 0). *ES* = effect size, calculated as the coefficient divided by the product of the standard error and square root of N. Approximate R^2 calculated as the fitted variance divided by the sum of the fitted, intercept, and residual variances. Model estimated using *R* lme4 and lmerTest packages.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 7

Linear Growth Model Results for Apex Participants' Contributing to Group Discussion

	Model 1					Model 2					Model 3					Model 4				
Fixed Effects	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*
Intercept (Baseline)	5.13	0.47	10.80	20 ***	1.00	4.95	0.49	10.18	23 ***	0.94	5.18	0.46	11.27	19 ***	1.04	5.07	0.49	10.38	25 ***	0.96
Gender	0.26	0.41	0.65	118	0.06	0.36	0.41	0.86	128	0.08	0.20	0.43	0.48	101	0.04	0.22	0.44	0.51	115	0.05
Returner	-0.54	0.32	-1.72	141	-0.16	-0.55	0.31	-1.77	142	-0.16	-0.41	0.34	-1.21	127	-0.11	-0.45	0.34	-1.32	128	-0.12
Mean Bunk Age	0.16	0.40	0.39	16	0.04	0.17	0.40	0.44	16	0.04	0.15	0.39	0.39	16	0.04	0.10	0.39	0.26	16	0.02
ASD						0.64	0.41	1.56	142	0.14						0.83	0.45	1.83	128	0.17
Comorbid						-0.97	0.45	-2.14	141 *	-0.20						-0.86	0.52	-1.67	127	-0.15
SRS-2 Score											-0.32	0.31	-1.02	130	-0.09	-0.29	0.35	-0.83	132	-0.08
Linear Growth	-0.02	0.01	-2.57	2024 *	-0.24	-0.02	0.01	-2.15	2024 *	-0.20	-0.02	0.01	-2.51	1834 *	-0.23	-0.02	0.01	-2.17	1834 *	-0.20
Gender*Time	0.00	0.01	0.58	2024	0.05	0.00	0.01	0.37	2024	0.03	0.01	0.01	0.81	1834	0.07	0.01	0.01	0.80	1834	0.07
Returner*Time	-0.01	0.01	-1.41	2025	-0.13	-0.01	0.01	-1.43	2025	-0.13	-0.01	0.01	-1.87	1834	-0.17	-0.01	0.01	-1.75	1834	-0.16
Mean Bunk Age*Time	0.01	0.01	1.71	2024	0.16	0.01	0.01	1.66	2024	0.15	0.02	0.01	2.27	1834 *	0.21	0.02	0.01	2.45	1834 *	0.23
ASD*Time						0.00	0.01	-0.54	2025	-0.05						-0.02	0.01	-2.06	1835 *	-0.19
Comorbid*Time						0.01	0.01	1.09	2024	0.10						0.02	0.01	1.60	1834	0.15
SRS-2 Score*Time											0.01	0.01	0.87	1835	0.08	0.01	0.01	0.81	1835	0.07
Random Effects	*Var*					*Var*					*Var*					*Var*				
Bunk Intercept (Baseline)	0.68					0.70					0.41					0.46				
Child Intercept (Baseline)	8.14					7.86					8.43					8.23				
Residual	7.18					7.18					7.36					7.34				
Model Fit																				
Approximate R^2	0.36					0.36					0.36					0.36				
BIC	10748.00					10773.80					9806.50					9830.30				
Deviance (-2LL)	10663.70					10658.80					9708.10					9701.60				
Residual *df*	2130					2126					1927					1923				
LRT Chi-square test	--					4.90					950.70 ***					6.50				

Note. $N = 117$. Campers measured at 19 timepoints; time coded in days from baseline (baseline = 0); SRS-2 = Social Responsiveness Scale, Second Edition; metrical predictors are standardized in z-scores and categorical predictors are effect coded. *ES* = effect size, calculated as the coefficient divided by the product of the standard error and square root of N. Approximate R^2 calculated as the fitted variance divided by the sum of the fitted, intercept, and residual variances. Model estimated using *R* lme4 and lmerTest packages.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 8

Linear Growth Model Results for Apex Participants' Novel Contribution

	Model 1					Model 2					Model 3					Model 4				
Fixed Effects	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*
Intercept (Baseline)	2.32	0.25	9.34	169 ***	0.86	2.17	0.25	8.56	173 ***	0.79	2.38	0.26	9.32	156 ***	0.86	2.29	0.27	8.38	158 ***	0.77
Gender	0.16	0.25	0.66	169	0.06	0.28	0.25	1.13	173	0.10	0.09	0.26	0.34	156	0.03	0.15	0.27	0.57	158	0.05
Returner	-0.41	0.21	-2.00	169 *	-0.18	-0.38	0.20	-1.92	173	-0.18	-0.36	0.22	-1.66	156	-0.15	-0.34	0.21	-1.60	157	-0.15
Mean Bunk Age	0.28	0.20	1.37	168	0.13	0.31	0.20	1.55	173	0.14	0.27	0.21	1.29	156	0.12	0.27	0.21	1.29	158	0.12
ASD						-0.10	0.26	-0.39	172	-0.04						0.02	0.29	0.08	157	0.01
Comorbid						-0.37	0.29	-1.29	172	-0.12						-0.26	0.33	-0.79	157	-0.07
SRS-2 Score											-0.35	0.20	-1.76	156	-0.16	-0.27	0.22	-1.21	158	-0.11
Linear Growth	0.07	0.01	11.47	2025 ***	1.06	0.07	0.01	10.29	2025 ***	0.95	0.07	0.01	11.08	1834 ***	1.02	0.07	0.01	9.95	1834 ***	0.92
Gender*Time	0.00	0.01	-0.72	2025	-0.07	0.00	0.01	-0.37	2024	-0.03	-0.01	0.01	-0.84	1834	-0.08	-0.01	0.01	-0.74	1834	-0.07
Returner*Time	0.00	0.01	0.65	2025	0.06	0.00	0.01	0.62	2025	0.06	0.00	0.01	0.04	1834	0.00	0.00	0.01	-0.05	1834	0.00
Mean Bunk Age*Time	0.02	0.01	2.94	2024 **	0.27	0.02	0.01	2.96	2024 **	0.27	0.02	0.01	3.77	1834 ***	0.35	0.02	0.01	3.56	1834 ***	0.33
ASD*Time						0.02	0.01	2.51	2025 *	0.23						0.02	0.01	2.15	1835 *	0.20
Comorbid*Time						-0.02	0.01	-3.16	2025 **	-0.29						-0.02	0.01	-1.93	1834	-0.18
SRS-2 Score*Time											-0.02	0.01	-3.08	1835 **	-0.28	-0.02	0.01	-2.63	1835 **	-0.24
Random Effects	*Var*					*Var*					*Var*					*Var*				
Bunk Intercept (Baseline)	0.00					0.00					0.00					0.00				
Child Intercept (Baseline)	3.22					2.96					3.08					2.98				
Residual	4.61					4.58					4.72					4.71				
Model Fit																				
Approximate R^2	0.32					0.33					0.34					0.34				
BIC	9738.10					9749.00					8885.60					8907.20				
Deviance (-2LL)	9653.70					9634.00					8787.20					8778.50				
Residual *df*	2130					2126					1927					1923				
LRT Chi-square test	--					19.70 ***					846.80 ***					8.70				

Note. $N = 117$. Campers measured at 19 timepoints; time coded in days from baseline (baseline = 0); SRS-2 = Social Responsiveness Scale, Second Edition; metrical predictors are standardized in z-scores and categorical predictors are effect coded. *ES* = effect size, calculated as the coefficient divided by the product of the standard error and square root of N. Approximate R^2 calculated as the fitted variance divided by the sum of the fitted, intercept, and residual variances. Model estimated using *R* lme4 and lmerTest packages.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 9

Linear Growth Model Results for Apex Participants' Ignoring a Negative Stimulus

	Model 1					Model 2					Model 3					Model 4				
Fixed Effects	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*
Intercept (Baseline)	4.22	0.94	4.50	13 ***	0.42	4.20	0.94	4.45	13 ***	0.41	4.20	0.92	4.58	13 ***	0.42	4.20	0.93	4.51	14 ***	0.42
Gender	-1.00	0.37	-2.74	731 **	-0.25	-0.98	0.37	-2.62	770 **	-0.24	-1.07	0.38	-2.81	1941 **	-0.26	-1.08	0.39	-2.73	1940 **	-0.25
Returner	-0.12	0.29	-0.42	1022	-0.04	-0.11	0.29	-0.37	1024	-0.03	-0.34	0.30	-1.12	1932	-0.10	-0.34	0.30	-1.12	1932	-0.10
Mean Bunk Age	-1.25	0.86	-1.45	12	-0.13	-1.24	0.86	-1.43	12	-0.13	-1.30	0.84	-1.53	12	-0.14	-1.30	0.85	-1.54	12	-0.14
ASD						-0.11	0.38	-0.30	1018	-0.03						0.12	0.40	0.31	1932	0.03
Comorbid						-0.03	0.41	-0.06	1017	-0.01						-0.12	0.46	-0.25	1932	-0.02
SRS-2 Score											-0.39	0.28	-1.39	1933	-0.13	-0.39	0.32	-1.22	1933	-0.11
Linear Growth	0.10	0.02	5.46	2032 ***	0.50	0.10	0.02	5.28	2031 ***	0.49	0.10	0.02	5.17	1931 ***	0.48	0.09	0.02	4.70	1931 ***	0.43
Gender*Time	0.07	0.02	3.89	2032 ***	0.36	0.07	0.02	3.69	2030 ***	0.34	0.07	0.02	3.65	1931 ***	0.34	0.07	0.02	3.57	1931 ***	0.33
Returner*Time	0.00	0.01	0.21	2033	0.02	0.00	0.01	0.21	2032	0.02	0.02	0.02	1.12	1931	0.10	0.02	0.02	1.17	1931	0.11
Mean Bunk Age*Time	-0.03	0.01	-1.87	2031	-0.17	-0.03	0.01	-1.88	2031	-0.17	-0.03	0.02	-1.94	1931	-0.18	-0.03	0.02	-1.89	1931	-0.17
ASD*Time						-0.01	0.02	-0.26	2035	-0.02						-0.01	0.02	-0.52	1931	-0.05
Comorbid*Time						0.01	0.02	0.45	2031	0.04						0.00	0.02	0.15	1931	0.01
SRS-2 Score*Time											0.02	0.01	1.21	1931	0.11	0.02	0.02	1.21	1931	0.11
Random Effects	*Var*					*Var*					*Var*					*Var*				
Bunk Intercept (Baseline)	7.50					7.50					7.06					7.10				
Child Intercept (Baseline)	0.05					0.02					0.00					0.00				
Residual	38.29					38.29					37.87					37.85				
Model Fit																				
Approximate R^2	0.22					0.22					0.22					0.22				
BIC	14010.60					14040.10					12696.30					12726.10				
Deviance (-2LL)	13926.30					13925.10					12597.90					12597.40				
Residual *df*	2131					2127					1928					1924				
LRT Chi-square test	--					1.20					1327.20 ***					0.50				

Note. $N = 117$. Campers measured at 19 timepoints; time coded in days from baseline (baseline = 0); SRS-2 = Social Responsiveness Scale, Second Edition; metrical predictors are standardized in *z*-scores and categorical predictors are effect coded. *ES* = effect size, calculated as the coefficient divided by the product of the standard error and square root of N. Approximate R^2 calculated as the fitted variance divided by the sum of the fitted, intercept, and residual variances. Model estimated using *R* lme4 and lmerTest packages.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 10

Linear Growth Model Results for Apex Participants' Intentional Aggression to Staff

	Model 1					Model 2					Model 3					Model 4				
Fixed Effects	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*
Intercept (Baseline)	0.28	0.22	1.28	325	0.12	0.40	0.23	1.79	335	0.17	0.29	0.24	1.19	292	0.11	0.45	0.26	1.76	302	0.16
Gender	0.22	0.22	1.03	325	0.10	0.14	0.22	0.66	336	0.06	0.21	0.24	0.88	293	0.08	0.11	0.25	0.46	303	0.04
Returner	-0.05	0.18	-0.27	324	-0.02	-0.05	0.18	-0.29	334	-0.03	-0.04	0.21	-0.18	293	-0.02	-0.03	0.20	-0.13	303	-0.01
Mean Bunk Age	-0.23	0.18	-1.28	323	-0.12	-0.24	0.18	-1.36	333	-0.13	-0.21	0.20	-1.02	293	-0.09	-0.21	0.20	-1.03	304	-0.10
ASD						-0.27	0.23	-1.17	333	-0.11						-0.39	0.27	-1.43	301	-0.13
Comorbid						0.56	0.26	2.17	330 *	0.20						0.68	0.29	2.30	298 *	0.21
CBCL Score											0.10	0.18	0.54	297	0.05	0.02	0.19	0.08	305	0.01
Linear Growth	0.00	0.01	-0.04	2028	0.00	0.00	0.01	-0.35	2028	-0.03	0.00	0.01	0.23	1822	0.02	0.00	0.01	-0.11	1822	-0.01
Gender*Time	0.00	0.01	-0.32	2028	-0.03	0.00	0.01	-0.07	2028	-0.01	0.00	0.01	-0.38	1821	-0.04	0.00	0.01	-0.15	1821	-0.01
Returner*Time	0.00	0.01	0.57	2028	0.05	0.00	0.01	0.63	2028	0.06	0.01	0.01	0.80	1822	0.07	0.01	0.01	0.78	1822	0.07
Mean Bunk Age*Time	0.00	0.01	-0.27	2028	-0.03	0.00	0.01	-0.20	2028	-0.02	-0.01	0.01	-0.64	1821	-0.06	0.00	0.01	-0.61	1821	-0.06
ASD*Time						0.00	0.01	-0.06	2029	-0.01						0.01	0.01	0.52	1823	0.05
Comorbid*Time						-0.01	0.01	-0.77	2028	-0.07						-0.01	0.01	-1.05	1822	-0.10
CBCL Score*Time											0.01	0.01	0.73	1822	0.07	0.01	0.01	0.86	1823	0.08
Random Effects	*Var*					*Var*					*Var*					*Var*				
Bunk Intercept (Baseline)	0.00					0.00					0.00					0.00				
Child Intercept (Baseline)	1.41					1.34					1.53					1.45				
Child Slope (Linear Growth)	0.00					0.00					0.00					0.00				
Residual	8.38					8.37					9.20					9.19				
Model Fit																				
Approximate R^2	0.11					0.11					0.11					0.11				
BIC	10888.20					10913.10					9982.10					10006.80				
Deviance (-2LL)	10796.20					10790.40					9876.20					9870.70				
Residual *df*	2130					2126					1910					1906				
LRT Chi-square test	--					5.80					914.20 ***					5.50				

Note. $N = 117$. Campers measured at 19 timepoints; time coded in days from baseline (baseline = 0); CBCL = Child Behavior Checklist, Externalizing Problems composite; metrical predictors are standardized in z-scores and categorical predictors are effect coded. *ES* = effect size, calculated as the coefficient divided by the product of the standard error and square root of N. Approximate R^2 calculated as the fitted variance divided by the sum of the fitted, intercept, and residual variances. Model estimated using *R* lme4 and lmerTest packages.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 11

Linear Growth Model Results for Apex Participants' Repeated Noncompliance

	Model 1					Model 2					Model 3					Model 4				
Fixed Effects	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*
Intercept (Baseline)	0.08	0.03	2.49	464 *	0.23	0.08	0.04	2.28	472 *	0.21	0.10	0.04	2.64	443 **	0.24	0.09	0.04	2.39	444 *	0.22
Gender	0.04	0.03	1.34	470	0.00	0.05	0.03	1.45	473	0.00	0.04	0.04	1.12	444	0.00	0.04	0.04	1.17	444	0.00
Returner	-0.03	0.03	-0.98	469	-0.09	-0.02	0.03	-0.87	470	-0.08	-0.02	0.03	-0.50	445	-0.05	-0.01	0.03	-0.46	444	-0.04
Mean Bunk Age	-0.05	0.03	-1.93	460	-0.18	-0.05	0.03	-1.84	469	-0.17	-0.06	0.03	-2.00	444 *	-0.18	-0.06	0.03	-1.88	446	-0.17
ASD						-0.05	0.04	-1.46	468	-0.14						-0.04	0.04	-0.97	441	-0.09
Comorbid						0.03	0.04	0.68	464	0.06						0.02	0.05	0.45	436	0.04
CBCL Score											0.06	0.03	2.07	451 *	0.19	0.05	0.03	1.75	447	0.16
Linear Growth	0.00	0.00	-0.37	2028	-0.03	0.00	0.00	-0.17	2027	-0.02	0.00	0.00	-0.54	1822	-0.05	0.00	0.00	-0.27	1821	-0.03
Gender*Time	0.00	0.00	-0.31	2028	-0.03	0.00	0.00	-0.51	2027	-0.05	0.00	0.00	-0.27	1821	-0.02	0.00	0.00	-0.48	1821	-0.04
Returner*Time	0.00	0.00	-0.08	2028	-0.01	0.00	0.00	-0.19	2028	0.00	0.00	0.00	-0.20	1822	0.00	0.00	0.00	-0.26	1822	0.00
Mean Bunk Age*Time	0.00	0.00	0.91	2027	0.08	0.00	0.00	0.80	2027	0.00	0.00	0.00	1.07	1821	0.10	0.00	0.00	0.88	1821	0.08
ASD*Time						0.00	0.00	1.36	2029	0.00						0.00	0.00	1.46	1823	0.13
Comorbid*Time						0.00	0.00	-0.39	2027	0.00						0.00	0.00	-0.41	1821	0.00
CBCL Score*Time											0.00	0.00	-0.41	1823	-0.04	0.00	0.00	-0.03	1823	0.00
Random Effects	*Var*					*Var*					*Var*					*Var*				
Bunk Intercept (Baseline)	0.00					0.00					0.00					0.00				
Child Intercept (Baseline)	0.02					0.02					0.02					0.02				
Child Slope (Linear Growth)	0.00					0.00					0.00					0.00				
Residual	0.26					0.26					0.28					0.27				
Model Fit																				
Approximate R^2	0.07					0.07					0.06					0.06				
BIC	3371.00					3399.00					3179.40					3207.00				
Deviance (-2LL)	3278.90					3276.30					3073.50					3070.90				
Residual *df*	2130					2126					1910					1906				
LRT Chi-square test	--					2.60					202.80					2.60				

Note. N = 117. Campers measured at 19 timepoints; time coded in days from baseline (baseline = 0); CBCL = Child Behavior Checklist, Externalizing Problems composite; metrical predictors are standardized in z-scores and categorical predictors are effect coded. *ES* = effect size, calculated as the coefficient divided by the product of the standard error and square root of N. Approximate R^2 calculated as the fitted variance divided by the sum of the fitted, intercept, and residual variances. Model estimated using *R* lme4 and lmerTest packages.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 12

Linear Growth Model Results for Apex Participants' Verbal Abuse

	Model 1					Model 2					Model 3					Model 4				
Fixed Effects	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*
Intercept (Baseline)	0.27	0.22	1.24	291	0.11	0.27	0.23	1.18	294	0.11	0.23	0.20	1.13	285	0.10	0.20	0.22	0.92	292	0.09
Gender	0.30	0.22	1.39	291	0.13	0.30	0.22	1.37	294	0.13	0.23	0.20	1.11	286	0.10	0.25	0.21	1.22	293	0.11
Returner	0.13	0.18	0.74	290	0.07	0.14	0.18	0.76	293	0.07	0.29	0.17	1.66	286	0.15	0.29	0.17	1.70	293	0.16
Mean Bunk Age	0.18	0.18	0.99	289	0.09	0.18	0.18	1.00	293	0.09	0.25	0.17	1.50	286	0.14	0.27	0.17	1.60	294	0.15
ASD						-0.06	0.24	-0.26	292	-0.02						-0.14	0.23	-0.61	291	-0.06
Comorbid						0.04	0.26	0.14	290	0.01						0.01	0.25	0.05	288	0.00
CBCL Score											0.42	0.15	2.69	290 **	0.25	0.39	0.16	2.46	295 *	0.23
Linear Growth	0.02	0.01	2.28	2027 *	0.21	0.02	0.01	2.67	2027 **	0.25	0.02	0.01	2.48	1821 *	0.23	0.03	0.01	3.09	1821 **	0.29
Gender*Time	0.01	0.01	0.95	2027	0.09	0.00	0.01	0.56	2027	0.05	0.01	0.01	1.16	1821	0.11	0.00	0.01	0.59	1821	0.05
Returner*Time	-0.01	0.01	-1.58	2027	-0.15	-0.01	0.01	-1.64	2027	-0.15	-0.01	0.01	-1.94	1821	-0.18	-0.01	0.01	-1.93	1821	-0.18
Mean Bunk Age*Time	-0.01	0.01	-1.13	2027	-0.10	-0.01	0.01	-1.22	2027	-0.11	-0.01	0.01	-0.97	1820	-0.09	-0.01	0.01	-1.06	1821	-0.10
ASD*Time						0.00	0.01	-0.49	2028	-0.05						-0.01	0.01	-0.81	1822	-0.07
Comorbid*Time						0.02	0.01	1.63	2027	0.15						0.02	0.01	2.22	1821 *	0.21
CBCL Score*Time											0.00	0.01	-0.13	1822	-0.01	0.00	0.01	-0.42	1822	-0.04
Random Effects	*Var*					*Var*					*Var*					*Var*				
Bunk Intercept (Baseline)	0.00					0.00					0.00					0.00				
Child Intercept (Baseline)	1.58					1.54					1.09					1.04				
Residual	7.73					7.72					6.35					6.34				
Model Fit																				
Approximate R^2	0.13					0.14					0.14					0.14				
BIC	10725.30					10751.00					9265.20					9286.40				
Deviance (-2LL)	10640.90					10635.90					9166.90					9157.80				
Residual *df*	2131					2127					1911					1907				
LRT Chi-square test	--					5.00					1469.00 ***					9.10				

Note. $N = 117$. Campers measured at 19 timepoints; time coded in days from baseline (baseline = 0); CBCL = Child Behavior Checklist, Externalizing Problems composite; metrical predictors are standardized in z-scores and categorical predictors are effect coded. *ES* = effect size, calculated as the coefficient divided by the product of the standard error and square root of N. Approximate R^2 calculated as the fitted variance divided by the sum of the fitted, intercept, and residual variances. Model estimated using *R* lme4 and lmerTest packages.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 13

Linear Growth Model Results for Apex Participants' Leaving the Activity Area

	Model 1					Model 2					Model 3					Model 4				
Fixed Effects	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*	*Coeff*	*SE*	*t*	*df*	*ES*
Intercept (Baseline)	0.49	0.18	2.75	442 **	0.25	0.54	0.19	2.93	455 **	0.27	0.63	0.20	3.19	413 **	0.29	0.72	0.21	3.44	422 ***	0.32
Gender	0.18	0.18	1.00	443	0.09	0.16	0.18	0.87	456	0.08	0.10	0.20	0.51	414	0.05	0.05	0.20	0.23	423	0.02
Returner	0.40	0.15	2.72	442 **	0.25	0.41	0.15	2.81	453 **	0.26	0.55	0.17	3.32	415 ***	0.31	0.56	0.17	3.39	423 ***	0.31
Mean Bunk Age	-0.24	0.15	-1.68	439	-0.15	-0.24	0.15	-1.66	452	-0.15	-0.33	0.16	-2.02	415 *	-0.19	-0.32	0.16	-1.98	424 *	-0.18
ASD						-0.38	0.19	-1.99	451 *	-0.18						-0.32	0.22	-1.44	420	-0.13
Comorbid						0.42	0.21	1.99	447 *	0.18						0.45	0.24	1.88	415	0.17
CBCL Score											0.42	0.15	2.82	421 **	0.26	0.36	0.15	2.31	426 *	0.21
Linear Growth	-0.01	0.01	-1.25	2025	-0.12	-0.01	0.01	-1.45	2025	-0.13	-0.01	0.01	-1.59	1819	-0.15	-0.02	0.01	-1.79	1819	-0.17
Gender*Time	0.00	0.01	0.15	2025	0.01	0.00	0.01	0.28	2025	0.03	0.00	0.01	0.45	1819	0.04	0.01	0.01	0.62	1819	0.06
Returner*Time	-0.01	0.01	-1.85	2026	-0.17	-0.01	0.01	-1.86	2026	-0.17	-0.02	0.01	-2.18	1819 *	-0.20	-0.02	0.01	-2.21	1819 *	-0.20
Mean Bunk Age*Time	0.01	0.01	1.73	2025	0.16	0.01	0.01	1.74	2025	0.16	0.01	0.01	2.02	1818 *	0.19	0.01	0.01	2.00	1819 *	0.18
ASD*Time						0.01	0.01	1.03	2027	0.09						0.01	0.01	0.79	1821	0.07
Comorbid*Time						-0.01	0.01	-1.31	2025	-0.12						-0.01	0.01	-1.18	1819	-0.11
CBCL Score*Time											-0.01	0.01	-1.55	1820	-0.14	-0.01	0.01	-1.27	1820	-0.12
Random Effects	*Var*					*Var*					*Var*					*Var*				
Bunk Intercept (Baseline)	0.00					0.00					0.00					0.00				
Child Intercept (Baseline)	0.66					0.63					0.66					0.64				
Child Slope (Linear Growth)	0.00					0.00					0.00					0.00				
Residual	7.06					7.05					7.63					7.62				
Model Fit																				
Approximate R^2	0.06					0.06					0.07					0.07				
BIC	10473.80					10499.20					9574.00					9600.50				
Deviance (-2LL)	10381.80					10376.40					9468.20					9464.40				
Residual *df*	2130					2126					1910					1906				
LRT Chi-square test	--					5.40					908.20 ***					3.80				

Note. $N = 117$. Campers measured at 19 timepoints; time coded in days from baseline (baseline = 0); CBCL = Child Behavior Checklist, Externalizing Problems composite; metrical predictors are standardized in z-scores and categorical predictors are effect coded. *ES* = effect size, calculated as the coefficient divided by the product of the standard error and square root of N. Approximate R^2 calculated as the fitted variance divided by the sum of the fitted, intercept, and residual variances. Model estimated using *R* lme4 and lmerTest packages.

* $p < .05$, ** $p < .01$, *** $p < .001$.

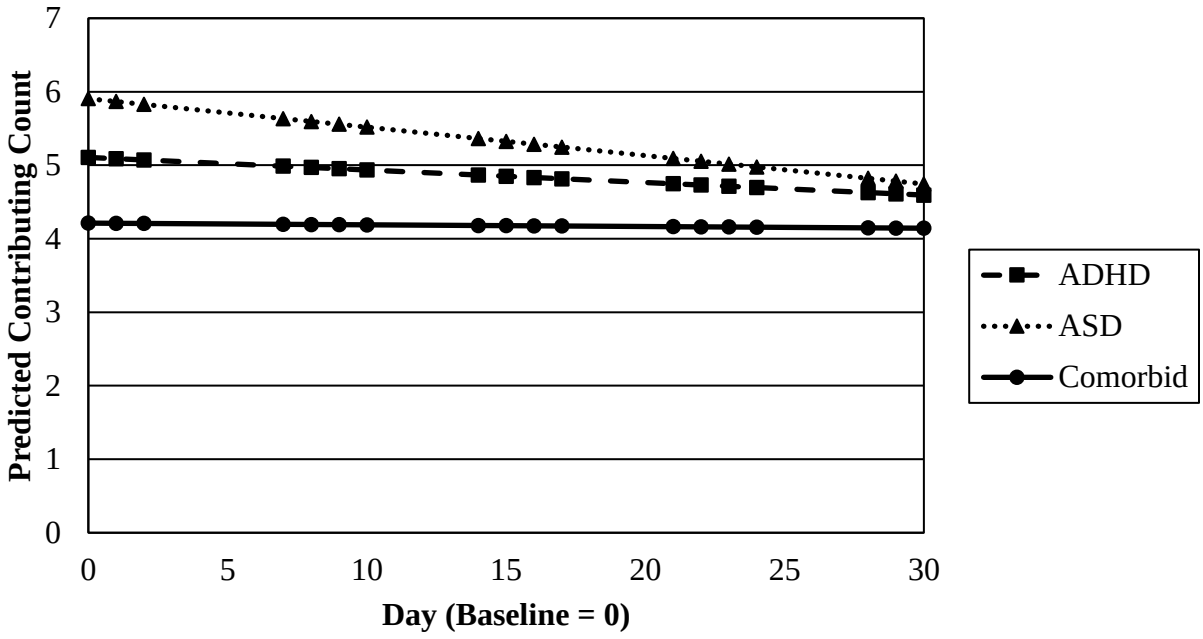
Figure 1

Representation of Apex Fidelity Check Procedures During a 60-minute Sports Activity

Note. Fidelity was monitored by evaluating one intervention component for equal intervals of a sports game or skills drill. For example, use of praise was evaluated for 15 minutes during Interval 2. The pre-activity discussion was the only component evaluated in Interval 1. Standardized attention questions were evaluated for 10 minutes of Interval 4 before post-activity discussion. Adapted from the STP intervention manual (Pelham et al., 2012).

Figure 2

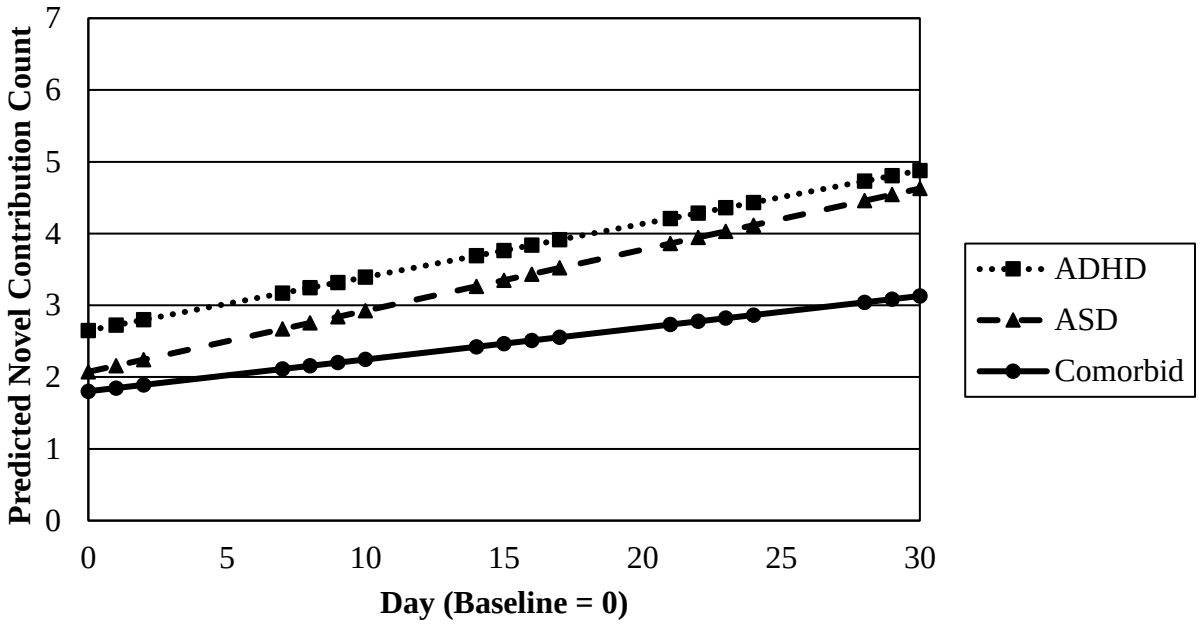
Predicted Contributing to Group Discussion Growth by Apex Participants' Diagnosis



Note. Figure constructed from Model 4 with SRS-2 score. Comorbid diagnosis significantly predicted Contributing to Group Discussion frequency at baseline before adding SRS-2 score, though campers with comorbid diagnoses were not predicted to demonstrate unique trajectories over time. However, campers with ASD were predicted to demonstrate a significant decrease in Contributing to Group Discussion over time after adding SRS-2 score to the model. SRS-2 = Social Responsiveness Scale, Second Edition.

Figure 3

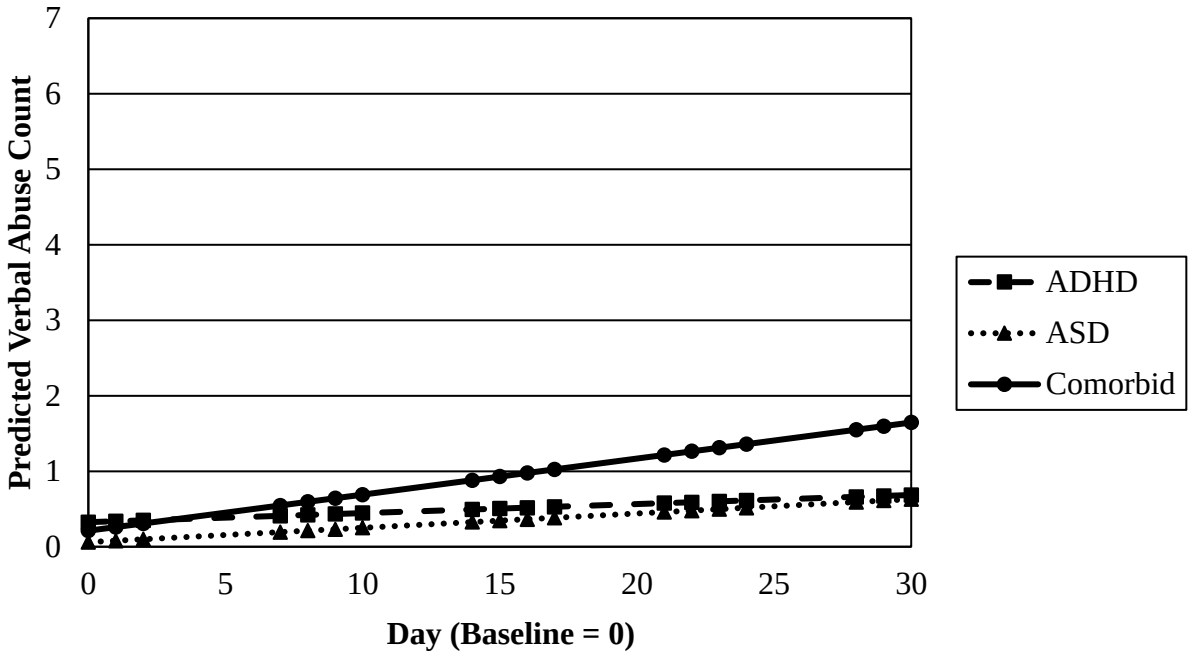
Predicted Novel Contribution Growth by Apex Participants' Diagnosis and Time



Note. Figure constructed from Model 2 without SRS-2 score. Diagnosis did not predict Novel Contributions at baseline. However, it significantly predicted growth over time for both campers with ASD and Comorbid diagnoses. Campers with both ASD and ADHD had a negative slope compared to campers with ASD alone who had a positive slope. Only ASD status remained significant after adding SRS-2 score to the model. SRS-2 = Social Responsiveness Scale, Second Edition.

Figure 4

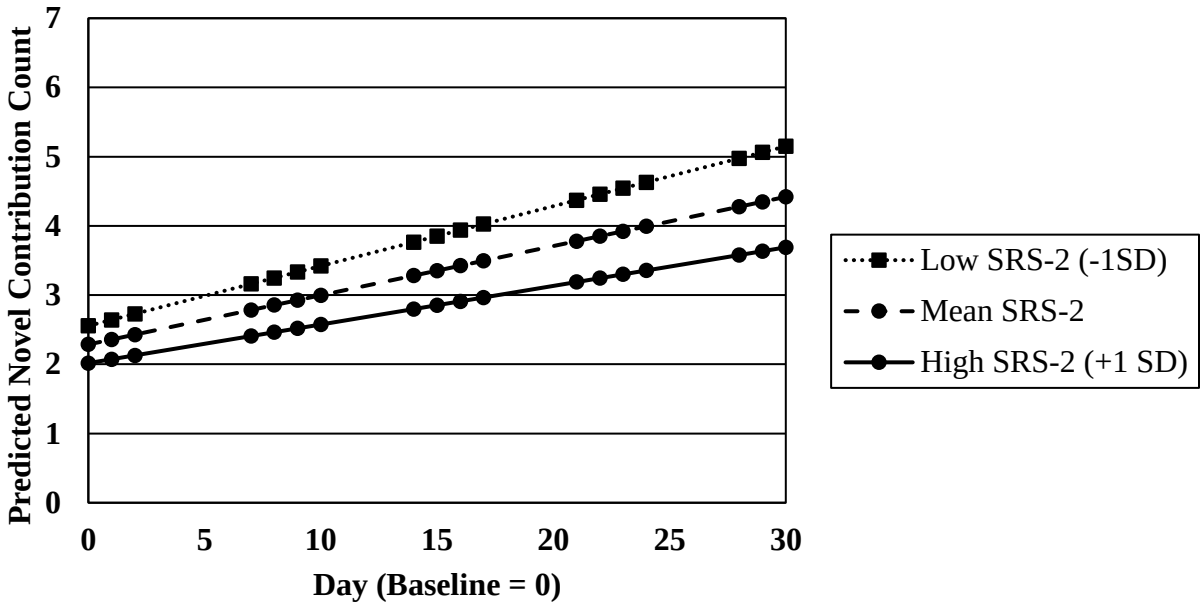
Predicted Verbal Abuse Growth by Apex Participants' Diagnosis and Time



Note. Figure constructed using Model 4 with CBCL scores. Diagnosis did not predict Verbal Abuse at baseline. However, it significantly predicted growth over time for campers with comorbid diagnoses after CBCL score was added to the model. CBCL = Child Behavior Checklist.

Figure 5

Predicted Novel Contribution Growth by Apex Participants' SRS-2 Baseline Score



Note. Figure constructed using Model 4 with diagnosis. The relationship between SRS-2 scores at baseline and Novel Contributions at baseline was nonsignificant, while SRS-2 scores significantly predicted growth in Novel Contributions over time. SRS-2 score remained a significant predictor both before and after adding diagnosis to the model. SRS-2 = Social Responsiveness Scale, Second Edition.