Title

Psychometric Properties of a Newly Developed Autism Screener: Accuracy and Predictive Factors in a Spanish-speaking Sample

Author

Christopher, Kourtney

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Psychometric Properties of a Newly Developed Autism Screener: Accuracy and Predictive Factors in a Spanish-speaking Sample

A dissertation submitted in partial
satisfaction of the requirements for the
degree Doctor of Philosophy in Education

by

Kourtney Christopher

2023

ABSTRACT OF THE DISSERTATION

Psychometric Properties of a Newly Developed Autism Screener:

Accuracy and Predictive Factors in a Spanish-speaking Sample

by

Kourtney Christopher

Doctor of Philosophy in Education

University of California, Los Angeles, 2023

Professor Catherine Lord Morrison, Chair

Few Level-2 screening instruments for autism cover a broad age range and include direct interactions with caregivers, and fewer have proven utility in Spanish-speaking populations. The study explored the psychometric properties of a newly developed tool in a large sample from South America (n = 295). The BOSA's sensitivity ranged from 0.46 to 1.00 and specificity from 0.44 to 0.93. Based on these findings, the BOSA demonstrates clinical utility to be used as a Level-2 screener in this population. Like other autism screening instruments, the Emotional and Behavioral Problems (EBPs) and differences across Autism Diagnostic Observation Schedule-Second Edition (ADOS-2) modules predicted false positive results on the screener. Potential considerations and implications are discussed.

The dissertation of Kourtney Christopher is approved.

Jeffrey J. Wood

Connie L. Kasari

So Hyun Kim

Catherine Lord Morrison, Committee Chair

University of California, Los Angeles

2023

Dedication

I dedicate this work to my family, who made it possible to move across the country to purse my dream of learning from the world's best, Dr. Lord. To my incredible sons, Kael and Kade Christopher, constantly reminding me, "You've got this," but in the same breath, "Are you done yet?" Both boys empower me to keep going and reach my goals. To my parents, Anita and Robin Christie, and David and Terri Nigh, who helped support and provide guidance over the years. To my aunt, Rita Martz, and grandmother, Peggy Rice, it's odd to say, but I am forever thankful for the COVID pandemic. I will never forget their willingness to take in my family when it felt like the world was collapsing around us. Moreover, my entire family's support, especially my mother homeschooling Kael, truly allowed me the capacity to focus on my studies and bring this to completion.

Lastly, I would like to extend special gratitude to Kael Christopher. He is very much embedded into my work and served as an anchor that guided my research questions in this study. I am incredibly thankful to be his mother and have the opportunity to learn from him daily. Kael's ideas, creations, and perceptions of the world are always fascinating and eye-opening. Through our trials and tribulations over the years, I believe I will be better suited to have a positive impact on those with autism spectrum disorder and their families.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Acknowledgments

I would especially like to thank Dr. Catherine Lord, my mentor. It was surreal when I met with her in October 2019, and she encouraged me to apply to her program. I still cannot believe I had the opportunity to learn from her. Dr. Lord is an incredible force in the autism field, with endless amounts of knowledge. Truly some of my best memories from graduate school were listening to her share her expertise and hearing her thought process. She has made such an impact on the field of autism, but most importantly, on the families she has served. Her commitment to patient-care is unmatched. I have also benefited from Dr. Lord's guidance with my son, and there is no better feeling as a parent to know you are in good hands. I hope to have this impact on families someday. Dr. Lord's dedication to my learning and persistence to keep pushing me to become a better researcher and clinician is immeasurable. I look forward to continuing to learn from her as she mentors me in my future endeavors.

I would also like to thank Drs. Kasari, Kim, and Wood, for their contributions to this project and participation on my committee. Their guidance through the process and feedback on this research have been invaluable. I want to thank Dr. Connie Kasari for her encouragement and guidance throughout my graduate studies. I enjoyed every class I took from her. I hope to be as thoughtful and dedicated to teaching future generations as she is. I also appreciate her willingness to add me to projects outside my comfort zone. The connections and knowledge I obtained in her studies will undoubtedly help me in the future. In addition, I am thankful for Dr. So Hyun Kim's willingness to take time and guide me professionally and personally. She is an incredible clinician, researcher, and amazing mother. She is truly someone I would love to emulate across all contexts. I am eternally grateful for her expertise and insights into this project, and I look

forward to future collaborations. Finally, I would like to thank Dr. Jeffery Wood for his invaluable insights into the scientific methodology of this work. I have benefited from his thoughtful comments and careful guidance on approaching research and answering questions effectively and effectively. I continue to use his intervention in my current work and hope to keep learning from him.

Much gratitude is extended to Dr. Christina Toolan for her guidance and organization of this project. Dr. Toolan is an incredible teacher, and I am blessed to have her unending personal and professional support. This project would have been difficult to complete without her careful attention to detail. I also want to thank Dr. Kyle Sterrett for his statistical consultations and feedback. I have enjoyed working with him over the years and hope to continue sharing ideas. In addition, I would like to acknowledge our South American colleagues who shared this data with our lab.

KOURTNEY CHRISTOPHER, M.ED. M.A

EDUCATIONAL BACKGROUND

University of California- Los Angeles June 2023
(Anticipated)
Doctorate in Education-Human Development and Psychology
Los Angeles, CA
Advisor: Catherine Lord, Ph.D.

University of California- Los Angeles August 2022
Master's in Education-Human Development and Psychology
Los Angeles, CA
Advisor: Catherine Lord, Ph.D.

University of Missouri- Columbia May 2009
Master's in Special Education (Emphasis in Autism Spectrum Disorder)
Columbia, MO
Advisor: Janine Stichter, Ph.D.

Avila University May 2007
Bachelor of Science in Biology (Minor in Chemistry)
Kansas City, MO
Advisor: *Stephen Daggett, Ph.D*

CLINICAL & RESEARCH EXPERIENCE

Global Autism Interactive Network (GAIN) Founder/Trainer March 2019- present
Weil Cornell- The Center for Autism and the Developing Brain
Activities: Attend monthly GAIN calibration meetings with the Autism Diagnostic Observation Schedule-Second Edition (ADOS-2) trainers, and serve as a panelist for webinar calibrations.

Leadership Education in Neurodevelopmental Disabilities (LEND) August 2022 – present
Predoctoral Fellow
Department of Pediatrics, Division of Developmental and Behavioral Sciences
University of Kansas Medical Center
Supervisors: Skylar Bellinger, Rene Jamison, Jamie McGovern, & Leni Swails, Jessica Schuttler, Ph.D.
Position: Psychology Predoctoral Fellow
Responsibilities: Provide assessment and intervention services, as well as diagnostic feedback under the supervision of a licensed psychologist as part of the LEND training program; formulate integrated reports in Epic; trained in Parent-Child Interaction Therapy (PCIT) and Girls Night Out (GNO); co-facilitated weekly GNO groups; conducted qualitative thematic analyses for childcare focus groups. Hours: 40 hours per week

UCLA Lord Lab September 2019 – August 2022
Supervisor: Catherine Lord, Ph.D.
Position: Graduate Student Researcher (GSR)
Responsibilities: Provide assessment services to adults under the supervision of a licensed psychologist as part of an ongoing longitudinal study looking at the development of autism; formulate integrated reports; participate in individual and group supervision; co-developing

modules for the Brief Observation of Social Commination Change (BOSCC); assisted in the development of new Level-2 screener, Brief Observation of Symptoms of Autism (BOSA); code videos for reliability and validation; manage and collect data for ongoing studies, Supported Play Research Outcomes Understanding Treatments (SPROUT), and Boosting Language Outcomes of Minimally Verbal Children with Autism Spectrum Disorder (BLOOM) under the directions of Drs. Lord, Kasari, and Tager-Flusberg; co-lead ADOS-2 clinical and research training across the country. Hours: Approximately 15-20 hours per week

MANUSCRIPTS IN PREPARATION AND UNDER REVIEW

Christopher, K., Elias, R., Sterrett, K., & Lord, C. (under review). A Longitudinal Study of Perceived Negative Impact in Parents of Individuals with Autism Spectrum Disorders from 2 to 25 years of age.

Christopher, K. & Lord, C. (in prep). Psychometric Properties of a Newly Developed Autism Screener: Accuracy and Predictive Factors in a Spanish-speaking Sample.

Schiltz, H., Clarke, E., Rosen, N., Gomez, S., Masjedi, N., **Christopher, K.,** & Lord, C. (under review). A Mixed-Methods Characterization of Family Support from Adolescence to Young Adulthood in Autism.

MANUSCRIPTS

Christopher, K. & Lord, C. (2022). Best Practice Assessments for Autism Spectrum Disorders in Schools. *Psychology in the Schools*. https://doi.org/10.1002/pits.22680

Nowell, K., **Christopher, K.,** & Sohl, K. (2020). ECHO Autism: Diagnostics: Equipping Community Psychologists and Practitioners to Deliver Best Practice Autism Spectrum Disorder Diagnoses. *Children's Health Care.* https://doi.org/10.1080/02739615.2020.1771564

Christopher, K., Bishop, S., Carpenter, L., Warren, Z., & Kanne, S. (2020) The Implications of Parent-Reported Emotional and Behavioral Problems on the Modified Checklist for Autism in Toddlers. *J Autism Dev Disord.* https://doi.org/10.1007/s10803-020-04469-5

NATIONAL/INTERNATIONAL PRESENTATIONS

ADOS/ADI-R Training

2023: 3 trainings: Kansas, New York, Pennsylvania, California
2022: 8 trainings: California, Arkansas, South Dakota, New York, Qatar, Missouri
2021: 4 trainings: California, Missouri, Kansas, and South Dakota (virtual)
2020: 3 trainings: California, Arkansas, and South Dakota (virtual)
2019: 11 trainings: California, Florida, Kentucky, Missouri, Nebraska, Ohio, Delaware
2018: 11 trainings: Arkansas, California, Kansas, Missouri, Nebraska, South Dakota
2017: 18 trainings: Arkansas, California, Kansas, Missouri, Texas
2016: 13 trainings: California, Colorado, Missouri, Nebraska, South Dakota, South Carolina
2015: 3 trainings: Missouri, Nebraska, South Dakota

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by core deficits in social communication and the presence of restricted repetitive behaviors. It is estimated to affect 1 in 44 North American children (Maenner et al., 2020). Among those diagnosed with ASD, there is broad phenotypic variability in presentation due to heterogeneity in core symptoms, cognitive and language abilities, adaptive skills, and co-occurring medical and psychiatric conditions (Lord et al., 2022). These heterogeneous presentations across the autism spectrum create barriers to the timely identification of children with ASD (Kanne & Bishop, 2021).

ASD can be reliably diagnosed in children as young as 15 months (Lord et al., 2006; Pierce et al., 2019), yet, the average age of diagnosis in the United States is 51 months (Maenner et al., 2020). These delays in diagnosis might have downstream effects, such as inhibiting timely access to evidence-based interventions that many believe lead to better long-term outcomes (Øien et al., 2021; Zwaigenbaum, Bauman, Choueiri, et al., 2015; Zwaigenbaum, Bauman, Stone, et al., 2015). As a result, researchers have begun to explore creative alternatives to expedite access to diagnostic evaluations (Zwaigenbaum, Bauman, Stone, et al., 2015; Zwaigenbaum & Warren, 2021) through the development of brief screening and diagnostic instruments. Some examples include the TELE-ASD-PEDS (Adiani et al., 2019; Corona et al., 2021) and Systematic Observation of Red Flags of ASD (SORF; (Dow et al., 2020). Both instruments assess for symptoms of autism in very young children, detecting autism symptomology through direct observations in addition to caregiver reports. While caregiver report instruments are often used to assess autistic symptoms, sensitivity, and specificity vary widely based on parental understanding and symptom presentation. Thus, direct observational

tools allow clinicians to observe and use their clinical judgment to decide whether the symptoms are attributable to autism.

A more recently developed instrument is the Brief Observation of Symptoms of Autism (BOSA; Lord et al., 2020; Dow et al., 2021). The BOSA was developed during the COVID-19 pandemic as a response to clinicians' and researchers' inability to complete the most widely used direct observation assessment, the Autism Diagnostic Observation Schedule-Second Edition (ADOS-2; Lord et al., 2012) due to personal protective equipment requirements. The BOSA uses direct observations to rate risk for ASD using ADOS-2 coding schemes and allows caregivers to interact directly with their child while the clinician observes from a socially distant location (i.e., behind a two-way mirror or through telehealth). Unlike many screening instruments, the BOSA covers a broad age range, and initial findings indicate strong psychometric properties (Dow et al., 2021). Therefore, beyond its initial intent to mitigate barriers put in place by the COVID-19 pandemic, the BOSA has the potential to address many of the challenges in the field related to timely access to diagnostic assessments for a broad range of individuals through brief, flexible, and direct observations.

These proposed studies will explore the psychometric properties of the BOSA in a Spanish-speaking international sample. Specifically, the utility of the BOSA as an autism screener will be examined to draw practical clinical implications for future use.

Identifying Autism Spectrum Disorder (ASD)

Diagnostic Criteria

ASD is a neurobiological condition typically diagnosed by highly trained licensed professionals, such as psychologists or physicians. Due to the heterogeneous nature of ASD, those diagnosing autism often need expertise in both typical child development and specific training in neurodevelopmental disorders. Most clinicians in the U.S. use the Diagnostic and Statistical Manual of Mental Disorders (DSM–5; American Psychiatric Association, 2013) to diagnose ASD. To meet DSM-5 diagnostic criteria for ASD, an individual must demonstrate or have demonstrated impairments in the two domains: difficulties in social communication and the presence of restricted, repetitive behaviors (RRB), either currently or by history. Within the social communication domain, an individual must demonstrate difficulties in three areas: 1) social-emotional reciprocity, 2) non-verbal communication, and 3) developing, maintaining, and understanding relationships. The RRB domain consists of four areas, but an individual only needs to exhibit behaviors in two of the four areas. These include 1) stereotyped or repetitive motor movements, use of objects, or speech, 2) insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behavior, 3) highly restricted, fixated interests that are abnormal in intensity or focus, and 4) hyper- or hypo-reactivity to sensory input or unusual interest in sensory aspects of the environment (APA, 2013).

The DSM-5 includes categories of behaviors and suggested disorders that have changed and evolved with time. ASD is an example of a disorder where the criteria have significantly changed and expanded over the years (Rosen et al., 2021). These changes have likely contributed to increased prevalence over the last several decades. Although important commonalities define ASD, the phenotypic expression of autistic core symptoms within an individual also changes

over time (Shulman et al., 2020). For these reasons, accurately assessing and identifying ASD across the lifespan is difficult.

To further complicate diagnostic decision-making, approximately 70% of children with ASD meet the criteria for at least one other mental health condition (Havdahl & Bishop, 2019). The most common co-occurring conditions include Attention Deficit Hyperactivity Disorder (ADHD), Oppositional Defiant Disorder (ODD), and Anxiety Disorders. These other mental health conditions, also termed Emotional and Behavioral Problems (EBPs), have symptoms that overlap with autism. Overlapping symptoms include challenges in maintaining attention, controlling aggression, and becoming overly concerned, worried, or fearful. These symptoms increase the complexity seen within autism and make diagnostic differentiation difficult.

ASD is commonly referred to as a heterogeneous disorder, requiring extensive training to identify and diagnose accurately. Finding a biomarker or gene that reliably predicts an autism diagnosis would be ideal as it would obviate the need for measures that rely only on behavioral phenotyping. For this reason, a tremendous amount of genetic and neuroscientific work has been attempting to pinpoint common genetic variants and biological markers to easily and quickly identify ASD (Geschwind & State, 2015; McPartland et al., 2020; Thapar & Rutter, 2021). Unfortunately, research has not identified a single genetic pathway or biomarker that always results in ASD (Chang et al., 2015; Chawner et al., 2019; Lord et al., 2018; Vorstman et al., 2017). Instead, phenotypic expression varies based on genotypic penetrance and other poorly understood factors, like epigenetics (Masini et al., 2020; Ramaswami & Geschwind, 2018; Thapar & Rutter, 2021). Moreover, hundreds of genetic anomalies are associated with ASD risk, and only about 20-30% of ASD can be explained through genetic findings (Chawner et al., 2019; Geschwind & State, 2015; Ramaswami & Geschwind, 2018; Rylaarsdam & Guemez-Gamboa,

2019). These include common and rare variants associated with the risk for autism (Lord et al., 2018). Biomarker researchers have noted the potential of quick and accurate identification of autism through eye-tracking and electrophysiological tests but have yet to scale their use in clinical settings (Lord et al., 2022; McPartland et al., 2020). Hence, there is a continued need for behavioral instruments to categorize autistic traits into discrete diagnostic categories efficiently and effectively.

Overview of Screening and Diagnostic Processes

There are multiple pathways for children to receive an ASD diagnosis, often beginning with caregivers noting concerns and seeking out professionals for further evaluation. Information collected by providers often includes screening questionnaires completed by caregivers to determine if a child is at risk for a wide range of developmental concerns. These population-based screeners are commonly referred to as level-1 screeners. The Ages and Stages Questionnaire-Third Edition (ASQ-3; Squires et al., 2009) is an example of a level-1 screener used to detect developmental concerns for young children in primary care settings (Brewer et al., 2020; Norris & Lecavalier, 2010). If it is determined that the child is at risk based on the level-1 screener, caregivers might then complete a level-2 screening instrument, which evaluates the risk for a specific disorder, like autism. Examples of level-2 screeners are the Social Communication Questionnaire (SCQ; Berument, 1999; Rutter et al., 2003) and the Social Responsiveness Scale–Second Edition (SRS-2; Constantino & Gruber, 2012), which are used to determine the risk for ASD. By indexing autism symptoms and risk, clinicians can quickly triage and refer families to appropriate clinics for diagnostic determination.

Depending on screening results, the child may be referred for a full diagnostic evaluation. During a diagnostic evaluation, clinicians directly assess the child and interview the caregivers

using various instruments. The evaluation frequently includes instruments quantifying autistic traits, like the ADOS-2 (Lord, Rutter, et al., 2012; Lord, Luyster, et al., 2012) and the Autism Diagnostic Interview-Revised (ADI-R; Rutter et al., 2003). In addition, clinicians assess the child's cognitive and language abilities through intelligence and developmental testing. Caregivers also complete questionnaires or interviews measuring their child's social, emotional, adaptive, and behavioral functioning in the home, which provides valuable information to aid in diagnostic determination. At the conclusion of the evaluation, clinicians make treatment recommendations based on their diagnostic impressions. For many disorders, including ASD, receiving the diagnosis opens pathways for recommended services and treatments.

ASD Screening Measures

Psychometrics

As noted above, the first step in identifying autism is often screening, using level-1 and level-2 screeners. Psychometrically, both screeners emphasize sensitivity, or the ability to identify those with the disorder, by casting a wide net. Sensitivity is contrasted with specificity, which is the ability of the instrument to correctly identify those who do not have the disorder. In general, for screening instruments to be considered psychometrically sound in the behavioral sciences, researchers have suggested that sensitivity and specificity be above 0.70 (Marks et al., 2008). Though both the sensitivity and specificity of a screener should be high, prioritizing sensitivity enables the instrument to capture any child who exhibits symptoms related to ASD and "flag" them for further diagnostic evaluations. Accordingly, children are less likely to be missed if they are at risk (high sensitivity), but more children will be "flagged" who do not have the disorder resulting in false positives (decreased specificity). In ASD and many disorders, prioritizing sensitivity over specificity in screening measures is considered acceptable, even

though caregivers might experience stress if their child is "flagged" but does not have the disorder (Petruccelli et al., 2022).

Whether a measure is designed to be a screener or for diagnosis, it is critical to consider the sample with which it was validated. Notably, the population in which the instrument was validated affects its accuracy in detecting the disorder. If the instrument is used in a population that is different from its validation sample, then its psychometric properties can be greatly impacted. For example, those referred to a clinic for a diagnostic question present with more clinically significant developmental and/or behavioral concerns (i.e., EBPs) compared to those in the general population (Havdahl & Bishop, 2019). If the instrument is not validated on a clinical sample, then the psychometrics based on the validation sample might not apply. Thus, clinicians should use instruments validated on populations that mirror patients in their clinical context to maximize accuracy.

There are many validated, age-based screeners for ASD; the most commonly used screeners are reviewed below.

Level-1 ASD Screening Instruments

For young children aged 12-30 months, The Modified Checklist for Autism in Toddlers-Revised with Follow-Up Interview (M-CHAT- R/F; Robins et al., 2014) is the most widely accepted and used as a population-based, level-1 screener. The M-CHAT- R/F is recommended by the American Academy of Pediatrics. Based on findings from the authors, it is effective as a population-based screener with strong sensitivity (0.85) and specificity (0.99) (Robins et al., 2014). However, findings from other population-based samples have been mixed, including alarmingly lower sensitivity (0.39; (Guthrie et al., 2019) and decreased specificity for minorities, those from low-income households, and Spanish speakers (Alonso-Esteban et al., 2020; Kimple

et al., 2014). In addition, low sensitivity (0.52) was found when using the M-CHAT in children who were born preterm (Kim et al., 2016). Decreased specificity (0.51) was also found when using the M-CHAT in samples of children referred to autism diagnostic clinics due to behavioral concerns potentially related to autism (Christopher et al., 2020; Lord et al., 2022; Øien et al., 2021). Some of these limits to sensitivity and specificity might be attributed to the presence of EBPs that overlap with ASD, which may contribute to positive screens for ASD (Christopher et al., 2020; Kim et al., 2016; Øien et al., 2021), negatively affecting the instrument's discriminative properties.

Level-2 ASD Screening Instruments

Caregiver Report Measures. The SCQ and SRS-2 are two of the most commonly used level-2 screeners. The SCQ is designed for children aged four years through adulthood and consists of yes/no questions derived from algorithm items from the ADI-R, a commonly used diagnostic instrument for ASD. The sensitivity of the SCQ ranges from 0.45 to 0.96, and specificity ranges from 0.54 to 0.95 depending on the sample (Chesnut et al., 2017; Norris & Lecavalier, 2010). Like the SCQ, the SRS-2 also characterizes ASD-related symptoms across a broad age range (i.e., 2.5 years through adulthood). Noted to be most efficiently used in general clinical and school settings (Constantino & Frazier, 2013), the SRS-2's sensitivity and specificity vary from 0.66-0.78 to 0.33-0.94, respectively (Constantino & Gruber, 2005; Kanne et al., 2018).

The Gilliam Autism Rating Scale-Third Edition (GARS-3; Gilliam, 2014) is often used in schools and is indicated for those aged 3-22. Based on the authors' data, the GARS-3's sensitivity and specificity ranged from 0.95-0.96 to 0.78-0.97, respectively (Samadi et al., 2022). However, the sensitivity of some screeners may vary based on the population assessed, especially with clinically referred samples (Montgomery et al., 2008; Norris & Lecavalier, 2010). For example,

though it performed well within its standardization sample, the GARS-3 demonstrated limited discriminative ability in referred samples (Camodeca, 2022). The Autism Spectrum Screening Questionnaire (ASSQ; (Ehlers et al., 1999) was developed to screen school-aged children for Asperger Syndrome. Based on previous findings, it performs well with sensitivities ranging from 0.62 to 0.91 and specificity from 0.86 to 0.91 (Ehlers et al., 1999; Posserud et al., 2009).

Though these screeners generally demonstrate strong sensitivity, several challenges remain with their use. All the aforementioned screeners are based on caregivers' perceptions of their child's behaviors and are thus subject to biases. For example, the SRS has been found to attribute autistic traits to individuals with EBPs inaccurately and those with profound intellectual disabilities (Gergoudis et al., 2020; Hus et al., 2013). Similarly, young children with EBPs and cognitive impairment are likely to meet ASD thresholds on the SCQ (Moody et al., 2017), even if they may not actually be autistic. Conversely, autistic youth exhibiting "milder" symptoms are likely not to meet clinically significant ASD thresholds on the SRS-2 and the ASSQ (Cederberg et al., 2018).

Direct Observation Measures. All screening instruments summarized above are questionnaires based on caregiver reports and do not include direct observations, which is an essential aspect of a diagnostic evaluation. Given this omission, they are not recommended for use as diagnostic instruments. Alternatively, the Childhood Autism Rating Scale (CARS-2; Schopler et al. 2010) combines caregiver reports and direct observation into its rating system. For this reason, many schools use the CARS-2 to index autistic symptomology as a diagnostic instrument. Recently, a novel approach focused on only using the CARS-2 direct observation scale rating produced sensitivity and specificity of 0.96 and 0.62, respectively (Sanchez & Constantino, 2020).

Innovative approaches have been developed to collect screening and diagnostic data outside the standard clinical context in the last ten years. Many of these approaches combine the caregiver's report with direct observations of the child within the home setting. The TELE-ASD-PEDS (Adiani et al., 2019; Corona et al., 2021) detects symptoms of ASD in preschool-age children through brief parent-child interactions. The Systematic Observation of Red Flags of ASD (SORF; (Dow et al., 2020) is used with toddlers, though it requires an hour-long observation to rate risk. Naturalistic Observation Diagnostic Assessment (NODA; Nazneen et al., 2015; Smith et al., 2017) is another direct assessment conducted at home that can detect risk in children up to seven years old. These innovative approaches provide an opportunity to collect information outside the standard clinical context through unbiased, direct observations. Consequently, these instruments might aid in expediting the referral process by triaging young children (under seven years old) to appropriate diagnostics clinics. Nevertheless, there remains a gap in the development of level-2 screeners collecting direct observations for older children and adults.

ASD Diagnostic Process

After an individual is screened and found at risk for autism, the next step is typically a comprehensive diagnostic evaluation. Licensed professionals using the DSM–5 can provide a formal, standardized diagnosis of ASD in clinical settings. Best practice guidelines and standards for diagnosing autism have been developed to ensure diagnostic accuracy (Aiello et al., 2017; Brian et al., 2019; Christopher & Lord, 2022; Lord et al., 2022; Mazurek et al., 2017). These guidelines almost uniformly suggest an evaluation that consists of indirect and direct assessments. Similarly, schools follow guidelines requiring data collection from multiple informants (e.g., caregivers, teachers, school psychologists) that include indirect and direct

assessments across a variety of settings to determine if a student's ability to access general education is negatively impacted by autism (Christopher & Lord, 2022). An educational classification of autism spectrum disorder is given in cases where a student meets the appropriate eligibility and requires additional support and services and is commonly tracked through an Individualized Education Plan with accompanying goals to support developmental and academic outcomes.

Indirect assessment of ASD symptomology includes data collected through interviews and questionnaires. Interviews can be conducted in person or via telehealth to obtain pertinent medical, family, and psychosocial history, with a specific emphasis on understanding the child's development and abilities (Aiello et al., 2017; Brian et al., 2019). Other questionnaires often included in diagnostic assessments inquire about a child's adaptive skills, behavioral profiles, and development to aid in diagnostic differentiation. Clinicians interpreting data from questionnaires should consider who completed the form, the context in which the reporter observes the child's functioning, and the reporter's reference group (Kanne et al., 2009). For example, a special education teacher who often observes children with significant behavioral challenges might not reach clinically significant thresholds when rating a child's behavior compared to a general education teacher whose reference group may not have the same degree of behavioral challenges. These considerations help clinicians interpret and contextualize the child's functioning across contexts to reach diagnostic clarity.

ASD Diagnostic Instruments

In contrast to screening instruments that emphasize sensitivity, diagnostic measures typically attempt to maximize both sensitivity and specificity to increase overall accuracy (Shreffler & Huecker, 2022). As noted, diagnostic assessments consist of both indirect and direct

instruments. The field of autism has widely accepted the use of two diagnostic instruments (e.g., ADI-R and ADOS-2) to be used in conjunction with one another and inform diagnostic decision-making (Christopher & Lord, 2022.; Esler & Ruble, 2015; Gray et al., 2014; Lebersfeld et al., 2021). The ADI-R is a standardized semi-structured caregiver interview designed to obtain detailed diagnostic information associated with ASD. The comprehensive interview is psychometrically sound, with sensitivities ranging from 0.75 to 0.82 and specificities from 0.72 to 0.85 when used in clinical and research contexts (Lebersfeld et al., 2021; Rutter et al., 2003).

The ADOS-2, on the other hand, is a direct observation tool administered by a trained clinician, usually within a clinical setting, to aid in diagnostic determination. Due to its strong psychometric properties, the ADOS-2 has risen to be regarded as a powerful tool for identifying autism across a broad range of ages and cognitive abilities (Shulman et al., 2020). The sensitivity of the ADOS-2 ranges from 0.83 to 0.92 and specificity from 0.81 to 0.94 when administered and scored by research-reliable clinicians (Lebersfeld et al., 2021; Shulman et al., 2020). Notably, the ADOS-2 validation sample included individuals clinically referred for a question of ASD. As previously discussed, these referred contexts include children with many symptoms from various disorders (not related to ASD), making diagnostic differentiation difficult, yet the ADOS-2 remained accurate (Lebersfeld et al., 2021; Shulman et al., 2020).

Cultural Considerations

The prevalence of ASD has increased over the past few decades and is reported to be between 0.5-2% worldwide (Chiarotti & Venerosi, 2020; Lord et al., 2018; Zablotsky et al., 2015). However, this reported prevalence is impacted by many factors, including access to care and the availability of culturally appropriate adaptations of screening and diagnostic instruments. For this reason, the identification of ASD differs by the region of the world in which the child

resides.

Disparities in access to care permeate high-income countries (HICs), like the U.S., as well as low-income and middle-income countries (LMICs) (Lord et al., 2022). Research in the U.S. suggests that the identification of ASD differs by social class, ethnicity, and race rather than actual prevalence (Elsabbagh et al., 2012; Fombonne, 2018; Maenner, 2021). More specifically, African American and Hispanic children are more likely to receive a diagnosis later than White children (Lopez et al., 2020; Maenner, 2021; Mandell et al., 2009). Alarmingly, this results in an average delay of 42.3 months for African American children between a caregiver's first concerns and the age of diagnosis (Constantino et al., 2020). Hispanic children average eight doctor visits before they receive a diagnosis (Lopez et al., 2020). Some of the delays in diagnoses can be attributed to cultural and socio-economic variability in caregiver expectations of behavior and development (Beacham et al., 2018; Blacher et al., 2014; Chesnut et al., 2017; Guthrie et al., 2019; Stevanovic et al., 2021; Stewart & Lee, 2017), and families mistrusting their medical providers (Moseley et al., 2007). Consequently, caregivers often report dissatisfaction and distress related to the diagnostic process (Crane et al., 2016; Goin-Kochel et al., 2006; Jacobs et al., 2020; Sanchez & Constantino, 2020; Zuckerman et al., 2015).

Another reason for disparities in access to care, especially internationally, is a paucity in the number of validated screening and diagnostic instruments in non-English speaking LMICs (Alonso-Esteban et al., 2020; Daley et al., 2013; Lord et al., 2022). Most of the recommended screening tools used to detect risk for ASD were developed and validated in HIC Western cultures. These screeners are often translated into different languages but are only occasionally validated in those languages. Validation studies are time-intensive, costly, and can be confounded by differences in dialect variations, such as in Spanish (Alonso-Esteban et al., 2020;

Kimple et al., 2014). A recent meta-analysis determined that only five screening tools have been used in Spanish-speaking communities to identify ASD (Alonso-Esteban et al., 2020). Four relied on caregiver reports of the five measures, which was problematic because Spanish-speaking caregivers reportedly experienced difficulty interpreting and understanding the questions (Alonso-Esteban et al., 2020; DuBay et al., 2021; Rea et al., 2019). In these situations, caregivers either skipped or incorrectly endorsed items, resulting in incorrect screening results for autism (DuBay et al., 2021; Kimple et al., 2014; Rea et al., 2019). Moreover, insufficient cross-cultural adaptations and translations also negatively impact the psychometric properties of instruments in Spanish-speaking samples (Alonso-Esteban et al., 2020; DuBay et al., 2021). In contrast, a level-2 screening instrument for young children that collected data through direct observations using a validated Spanish version (Hedley et al., 2015) of the Autism Detection in Early Childhood (ADEC; Young, 2007) was less vulnerable to cultural and linguistic differences (Alonso-Esteban et al., 2020). Thus, further work is needed to develop a brief, direct observation screening tools for ASD that are not impacted by cultural differences (Carruthers et al., 2018).

Diagnostic instruments are also plagued with similar but possibly even more profound difficulties due to the increased scope and cost of validation work. For example, although the "gold standard" diagnostic instrument, the ADOS-2, has been translated into over twenty languages, it has only been validated in Polish, German, Korean, and Portuguese (Chojnicka & Pisula, 2017; Lee et al., 2019; Medda et al., 2019; Pacífico et al., 2019). Determining the cross-cultural performance of an instrument requires the recruitment of large samples and clinicians trained in the appropriate use of diagnostic measures. Consequently, such studies are often scarce, underfunded, and underpowered to determine the instrument's diagnostic accuracy in diverse samples. Moreover, in the studies conducted, diagnostic accuracy results vary based on

the diagnostic instrument investigated (i.e., ADOS or CARS) and the language spoken (Chojnicka & Pisula, 2017; Medda et al., 2019; Overton et al., 2008; Pacífico et al., 2019; L. Smith et al., 2017; Stevanovic et al., 2021). For example, the ADOS and ADOS-2 performed well in German, Korean, and Polish-speaking populations. However, the CARS were not found to provide a valid cross-culture assessment for ASD when examining children from India, Jamaica, Mexico, and Spain (Stevanovic et al., 2021). Moreover, Harrison et al. (2017) found item-level differences (i.e., lack of measurement invariance) for eye contact, stereotyped language, and echolalia based on race and ethnicity rather than actual cross-cultural variations in ASD symptom expression. In that study, there was an increased likelihood for Black and Hispanic children to be rated higher on those items and inaccurately meet diagnostic thresholds (Harrison et al., 2017). Though this study was conducted in the U.S., it highlights cultural issues that may be even more pervasive when using measures across cultures. These important cultural factors need to be better understood, specifically in LMICs and non-English speaking countries, yet relevant research is limited.

Brief Observation of Symptoms of Autism (BOSA)

As noted earlier, the BOSA was developed during the COVID-19 pandemic in direct response to the need to have a substitute for the ADOS-2. The BOSA is a brief, play-based observation conducted by a caregiver to characterize autism symptomology (Dow et al., 2021). It can be administered in various contexts (i.e., home or clinic) across a broad age range of individuals (from toddlers through adults). Clinicians observe the interaction and later score behaviors using an ADOS-2 protocol to assess autism symptomology. It is only recommended to be scored by those trained and well-versed in the ADOS-2 (Dow et al., 2021). At the time of development, the BOSA was created to maximize sensitivity over specificity and used a

dichotomous rating system (i.e., presence [1] or absence of behaviors for items [0]). Results from the BOSA validation study in an English-speaking sample in the U.S. demonstrated high sensitivity and specificity, ranging from 0.86-0.98 and 0.70-1.00, respectively (Dow et al., 2021). These results are more commonly seen in diagnostic instruments rather than screeners. The authors noted several possible clinical implications for the BOSA, including its use as a screener. More specifically, the BOSA might be used as a level-2 screener to determine if a child needs to undergo a full diagnostic assessment (Dow et al., 2021). However, before further clinical recommendations are given, the BOSA should be validated in other samples.

While the BOSA demonstrated strong psychometric properties within its validation sample, there is a need to replicate its performance in other, more culturally and linguistically diverse samples. The initial validation sample consisted of primarily White, non-Hispanic, English-speaking, autistic individuals. Findings from other level-2 screening and diagnostic instruments note that cultural and linguistic differences affect discriminative thresholds for identifying ASD (Al Maskari et al., 2018; Beacham et al., 2018; Rea et al., 2019). Thus, it is essential to examine the validity of the BOSA in such samples. Next, the validation sample was exclusively rated by ADOS-2 "research reliable" clinicians (Dow et al., 2021). This term is often used in research studies and refers to clinicians who reach a pre-specified global standard of item-level reliability on the ADOS-2. Most clinicians trained on the ADOS-2 will not reach item-level reliability, and research has shown that clinicians who are not research reliable are less likely to rate individual items accurately (Kamp-Becker et al., 2018; Zander et al., 2016). The BOSA uses a limited number of specific items from the ADOS-2, which collapse into a binary schema of presence or absence to indicate concern for ASD. Thus, differences in item-level ratings on the ADOS-2 codes might have an even greater effect on the BOSA total scores

and discriminative thresholds. Therefore, exploring the BOSA's performance with individuals with less training is vital.

Another difference is related to who administers the BOSA to the child. The BOSA was developed during the COVID-19 pandemic to respond to clinicians' inability to complete direct observations and interactions with patients. As a result, it can be administered by caregivers or partners while clinicians watch from a socially distant location. However, research assistants, not caregivers, directly interacting with the participants completed a majority of the BOSAs in the validation sample. Dow and colleagues (2021) noted that clinicians trained on the ADOS-2 intentionally change their behaviors to provide opportunities to assess better the presence or absence of ASD-specific behavior. These differences in the interactant's style of administration might impact the instrument's ability to provide a rich sample of autism symptoms in a short period of time (Dow et al., 2021). That is, in Dow et al. (2021), with the interactant being research assistants (even though they were not ADOS trained), there may be differences in the types of behaviors elicited by these examiners compared to caregiver interactions. Thus, future work should explore if the psychometric properties of the BOSA remain intact when administered by caregivers.

Lastly, ASD is a complex disorder that involves diagnostic differentiation and often presents with co-occurring disorders (Havdahl & Bishop, 2019). Co-occurring disorders in ASD complicate clinicians' diagnostic certainty and accuracy. As noted, research has shown that EBPs impact discriminative thresholds of screening and diagnostic measures (i.e., decreased specificity) by increasing the likelihood of false positives (Christopher et al., 2020; Georgiades et al., 2011; Havdahl et al., 2016; Hus et al., 2013). Researchers have not studied the impact of these EBPs on the BOSA's psychometric properties, which may be critical to understanding its

accuracy and performance.

The Study

The prevalence of ASD is rising worldwide. Prompt and accurate identification of ASD is essential to open pathways for children to engage in evidence-based interventions (Zwaigenbaum et al., 2015). There is a critical need for an ASD screening instrument that is validated in different cultural contexts and includes direct observations across a broad age range. Given that Spanish is one of the most commonly spoken languages worldwide (Cantor-Cutiva et al., 2021) and is the second most widely spoken language in the U.S. (Cantor-Cutiva et al., 2021), targeted efforts to determine a reliable and valid level-2 screening instrument in Spanish-speaking communities would be of great benefit. For these reasons, I explored the utility of the BOSA as a level-2 screener in a Spanish-speaking sample from South America. This promising new tool has the potential to address several challenges beyond its initial intent of addressing a diagnostic need in response to the COVID-19 pandemic.

My dissertation evaluated the usability and utility of the BOSA as a level-2 screening instrument through rigorous psychometric investigations using a Spanish-speaking sample of individuals with and without ASD through the following two aims.

Aim 1. Determine the psychometric properties of the BOSA (i.e., sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy) in a Spanish-speaking clinical sample.

Aim 2. Explore predictors (e.g., EBPs, age, ADOS module) of false positives and negatives in the BOSA in this sample.

Methods

Procedures

The dataset includes a large international sample of Spanish-speaking participants from South America. A free, web-based BOSA training was provided to clinicians at the onset of the COVID-19 pandemic as providers searched for an alternative to the ADOS-2. To aid in the development of the BOSA, clinicians who attended the training, including community practitioners and researchers, were asked to share BOSA data of individuals who were seen for clinical and research purposes. The sample consists of de-identified data shared by South American clinicians through Research Electronic Data Capture (REDCap; Harris et al., 2009, 2019).

Participants

The entire sample consisted of 295 unique participants ranging from age 1.33 to 49.33 years. Participants were either seen for diagnostic determination or as controls with no diagnoses as part of a research study. Table 1 depicts the number of BOSA observations, while Table 2 includes participant BOSA data across the five ADOS-2 Modules. The Minimally Verbal BOSA is coded using either using an ADOS-2 Toddler or Module 1 protocol, depending on the child's age. The remaining BOSA versions correspond to the ADOS-2 based on language and age (i.e., Phrase Speech Young Fluent uses either Module 2 or Module 3, Fluent Speech 1 uses Module 3, and Fluent Speech 2 uses Module 4). To replicate the findings from Dow et al. (2021), BOSA data will now be discussed by ADOS-2 Module. The appropriate ADOS-2 module is determined by the age and language level of the participant. The Toddler Module is designed for young children at or under 30 months, while Module 1 is used for children at least 31 months of age with minimal speech. Module 2 is commonly used for young children with short phrase speech.

Both Modules 3 and 4 are for individuals with fluent speech, and Module 4 is used with adults with responsibilities (i.e., a job and/or over 16 years old).

Participant Variables

Demographics. The participant sample is predominately male (69.80%), reflecting a similar male-to-female ratio in autism diagnoses more generally (Loomes et al., 2017). The sample was mainly White (84.75%) and Hispanic (99.70%). All BOSAs were conducted in Spanish in Argentina or Chile.

Diagnosis. The clinical diagnosis was determined by the reporting clinician. Eight categories of diagnoses were created. For those with ASD, there are two categories, (1) ASD only and (2) ASD with Intellectual and Developmental Disorder (IDD)/Global Developmental Delay (GDD). Those not diagnosed with ASD presented with a broad range of other disorders, including Language Disorders, Attention Deficit Hyperactivity Disorder (ADHD), IDD/GDD, Anxiety/Depression, and a category labeled "Other" that includes various genetic and social-emotional diagnoses. Lastly, the final category has been termed "No diagnosis" (38.60%), which included typically developing controls. 33.56% of the sample received an ASD diagnosis. Given the large proportion of participants who ultimately did *not* have ASD, this variation increases heterogeneity and diagnostic complexity. See Table 3 for further participant data.

Other Testing. In some cases, assessment and diagnostic data (i.e., intelligence testing [34.58%], autism interviews [14.58%; ADI-R], and adaptive skills [51.86%]) were reported. Intelligence testing consisted of developmental and cognitive testing with accompanying standard scores. Adaptive behaviors were collected using the Vineland Scales of Adaptive Behaviors (VABS; Sparrow et al., 2016). The Adaptive Behavior Composites (ABC) from the VABS were used to reflect the participants' adaptive functioning. Total scores from the ADI-R

caregiver-reported autism interview were compiled to determine if the participant exceeded the recommended cutoffs associated with ASD on the diagnostic algorithm across all domains (e.g., A-D). See Table 4 for further details on assessments broken down by ADOS-2 modules.

Clinician Variables

Reliability Status. In the current sample, clinicians reported their level of reliability on the ADOS-2. There are two levels of reliability that clinicians can report: Research-Reliable (RR), which in the United States means that they have reached 80% item level agreement with an ADOS-2 trainer or another research-reliable ADOS-2 coder across the ADOS-2 modules, or Clinically Trained (CT), which means they have attended an ADOS-2 training. 24.07% of clinicians reported that they are research reliable, while the remaining 75.93% are clinically trained. For the purpose of this study, I combined CT and RR clinicians for analyses unless otherwise stated.

Profession. Clinicians reported their occupations, which consisted of child neurologists, psychologists, and speech and language pathologists. Clinician occupational data is presented in Table 5.

Best Estimate Diagnosis. Based on direct and indirect assessments, each clinician determined the best estimate diagnosis. In all cases, the BOSA was used as the direct observation tool, in some cases, supplemented with information from the ADI-R, Vineland Scales of Adaptive Behaviors, and cognitive or developmental assessments. The final diagnostic determination was shared via REDcap.

Measures

***Brief Observation of Symptoms of Autism*.** The Brief Observation of Symptoms of Autism (BOSA; Lord et al., 2020) was developed during the COVID-19 pandemic to provide a

naturalistic social context with standardized materials and activities for caregivers to interact with their children. It has four versions, each with a slightly different set of instructions: Minimally Verbal [MV], Phrase Speech and Young Fluent [PSYF], Fluent Speech for older children/early adolescents [F1], and Fluent Speech for older adolescents/adults [F2]. Clinicians observe interactions from a socially distant location (i.e., behind a two-way mirror, in-person from 6 feet away, or through a video platform) and score the interactions using an ADOS-2 protocol that corresponds to the age and language level of the participant (e.g., Toddler or Modules 1-4). Scores on ADOS-2 items range from 0 to 3, where zero indicates that the behavior is not consistent with ASD, and three strongly indicates that the behavior is sufficiently present to interfere with other behaviors on the ADOS. To minimize over-interpretation of these brief observations (Dow et al., 2021), the ADOS-2 item scores are collapsed into binary BOSA scores (1=presence of behavior, 0=not observed). BOSA items are summed across two core autism domains, assessing autism core features (i.e., social communication; restricted and repetitive behaviors) to create a total score to determine if the child meets the cutoffs associated with ASD risk (Dow et al., 2021). See Table 6 for the BOSA cutoff and ranges of risk from Dow and colleagues. Appendix A contains the BOSA DSM-5 checklist with ADOS-2 score conversions and associated cutoff scores by module. Data for one algorithm code (e.g., descriptive gestures) were unavailable; thus, the total cutoff score was decreased by a point to 8. Ranges of concerns were also reduced by one point across categories (e.g., 0-5 Little-to-No, 6-7 Mild-to-Moderate, and 8+ Moderate-to-Severe).

Someone familiar (i.e., BOSA interactants) with the participant completed the BOSA most often in clinical contexts (70.14%). The BOSA interactants included caregivers (biological, step, and adoptive), grandparents, partners, siblings, and aunts.

Emotional Behavioral Problems (EBPs). Following Havdahl et al.'s (2016) work with the ADOS-2, a total EBP score was created by summing all the "other behavior items" on the BOSA, ranging from 0 to 3 for Modules 1-4 and 0-4 for the Toddler Module. These items included Overactivity, Anxiety, Irritability/Fussiness, and Tantrums, Aggression/Negative or Disruptive Behavior.

Analytic Plan

Aim 1. Validate the BOSA in a Spanish-speaking Sample

This study replicated analyses conducted by Dow et al. (2021) to determine the sensitivity, specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and accuracy of the BOSA by module.

Sensitivity and specificity have been previously discussed as indicators of the instrument's ability to accurately identify those who have and do not have the disorder. The PPV refers to the likelihood of having a disorder in the context of a positive test result, while NPV is the likelihood of having a disorder in the context of a negative test result (Carvajal & Rowe, 2010). Higher rates of false positives affect PPV and specificity, which are likely to occur in referred contexts. Accuracy is when the test correctly identifies the presence or absence of a disorder.

Power analysis guidelines from Bujang & Adnan (2016) were used to determine recommended sample sizes to evaluate sensitivity and specificity. Based on those estimates assuming the prevalence rate is 50%, with a Type-1 error of under 5%, a minimum sample size of 40 participants (including 20 with ASD) within each ADOS module will be required to achieve a minimum power of 80% (actual power=80.4%) for detecting a change in the percentage value of sensitivity and specificity of a screening test from 0.50 to 0.80, based on a

target significance level of 0.05 (actual p=0.039) (Bujang & Adnan, 2016). Table 1 depicts the sample size by diagnosis and module.

The psychometric properties of the BOSA were examined by determining the number of false positives, false negatives, true positives, and true negatives by comparing the final diagnosis to recommended BOSA cutoffs by version. If an instrument's scores incorrectly indicated that the child met the criteria for a disorder, it was considered a false positive. False negatives resulted when an instrument incorrectly indicated a child did not meet the designated threshold, despite the child having ASD. The final two categories, true positives, and negatives, demonstrated that the instrument accurately categorized the child's behaviors. The following formulas were used to calculate sensitivity, specificity, PPV, NPV, and accuracy.

Sensitivity = True Positives/(True Positives +False Negatives)

Specificity = True Negatives/(True Negatives +False Positives))

Positive Predictive Value = True Positives/(True Positives +False Positives)

Negative Predictive Value = True Negatives/(True Negatives +False Negatives)

Accuracy = (True positives + True Negatives)/(True Positives + False Positives + False Negatives + True Negatives)

Aim 2: Explore Predictors of False Positives and Negatives

For the second aim, two separate logistic regressions were conducted to determine predictors. The first logistic regression examined false positives as the dependent variable with EBPs (continuous), participant's age (continuous), and ADOS-2 module (categorical) as the independent variables. The second logistic regression used false negatives as the dependent variable with EBPs, participant's age, and ADOS-2 module as the independent variables.

Expected Results

The ADOS-2 has been translated into Spanish but has not yet been validated. Prior research has noted item-level differences in the ADOS-2 items based on the ethnicity of the child (i.e., Hispanic) Harrison et al., 2017, and the BOSA results are dichotomized from item-level ADOS-2 codes (Dow et al., 2021). Thus, it is hypothesized that BOSAs given in Spanish will result in an increased likelihood of false positives and negatives, thus reducing overall accuracy.

Related to the second aim of the study, findings from the ADOS-2 demonstrate that the presence of EBPs increases the likelihood of false positives (Greene et al., 2021; Havdahl et al., 2016b; Molloy et al., 2011). Thus, I expected a primary predictor to be increased numbers of EBPs, resulting in false positives on the BOSA. Based on past research demonstrating that children with more complex speech are more likely to result in false negatives on other direct observation measures, I anticipate the ADOS-2 module to be predictive of false negatives. More specifically, children given Modules 2 or 3 will be more likely to be missed on the BOSA when compared to Modules 1 and Toddler.

Results

Aim 1: Validate the BOSA in a Spanish-speaking Sample

For most of the ADOS-2 modules, the BOSA demonstrated higher sensitivity relative to specificity when using the published cutoff scores (See Table 7). The Toddler Module's sensitivity was 82%, specificity was 47%, and accuracy was 60% using the recommended cutoff of 6. The PPV was 47%, while NPV was 82%. Ninety-four percent of the ASD group scored as mild-to-moderate or moderate-to-severe concern. The Module 1 sample resulted in similar findings with a sensitivity of 70%, specificity of 44%, accuracy of 54%, PPV of 41%, and NVP of 74% at a recommended cutoff of 5. Seventy percent of participants with ASD scored as mild-to-moderate or moderate-to-severe concern. The Module 2 sample consisted of many false

negatives and few true positives, yielding reduced sensitivity. Module 2 scoring resulted in a sensitivity of 46%, specificity of 93%, and accuracy of 79%. PPV and NVP remained high at 75% and 90%, respectively. Only about half (54%) of ASD participants scored in ranges of mild-to-moderate or moderate-to-severe concern using Module 2 scoring. Sensitivity for Module 3 was 89%, specificity was 79%, accuracy was 83%, PPV was 67%, and NPV was 94%, using the recommended cutoff of 6. All ASD participants scored as mild-to-moderate or moderate-to-severe concern using the Module 3 scoring. Finally, the Module 4 sample size was insufficient, and analyses and findings were treated as exploratory. Similar to Module 3, Module 4 demonstrated high sensitivity (100%), specificity (73%), accuracy 81%, PPV (64%), and NVP (100%), using a cutoff of 3. All participants with ASD scored in the Moderate-to-Severe range of concern. Overall, Modules 2, 3, and 4 demonstrated high specificity. This finding might be related to a high number of true negatives having no other diagnoses (i.e., typically developing), 86%, 84%, and 75%, respectively. This high rate of individuals not diagnosed with disorders aside from ASD is unlikely to occur in referred clinical contexts. Thus, conclusions related to high accuracy should be interpreted with caution.

Additional exploratory analyses were conducted to examine the BOSAs' psychometric properties when rated by Clinically Trained (CT) and Research Reliable (RR) clinicians. Overall, CT clinicians demonstrated lower accuracy (M = 60.30) compared to RR clinicians (M = 84.50). A one-way ANOVA revealed that there was a statistically significant difference in accuracy between the groups (F(1, 293) = [14.747], p = <.001). Due to limited sample sizes within modules, further comparative analyses were not able to be completed. Instead, Table 8 describes the psychometric properties of each ADOS module by training level.

Aim 2: Explore Predictors of False Positives and Negatives

Analyses were carried out using the Statistical Package for Social Sciences (SPSS) version 29. A logistic regression was conducted to determine which variables predicted false positive results on the BOSA with a significance level set at alpha = 0.05 (two-tailed). See Table 9 for the results. The presence of Emotional and Behavioral Problems (EBPs) significantly predicted the BOSA resulting in a false positive (β= .422, p < .01). More specifically, the relative odds (odds ratio) indicated that with a 1-point increase in EBPs, there was a 1.525 greater chance of a false positive. ADOS-2 module also significantly predicted false positives ($p < .01$). False positives were more likely to occur in the modules used with younger children (i.e., Toddler and Module 1) than with Modules 2 and 3 ($p < .01$). Age was not a significant predictor (β= -.003, $p < .52$) of a false positive result. For further details on false positives across ADOS-2 modules, see Table 11.

A separate logistic regression analysis was conducted to determine which variables predicted false negatives on the BOSA. Table 12 depicts the frequencies of false and true negatives counts by module. Results from the logistic regression did not reveal any significant predictors (age:β= .004, $p < .68$; ADOS-2 Module: $p < .30$; or EBPs: β= .286, $p < .16$), see Table 10 for results.

Item-level analyses were conducted to better understand false negative results for Modules 1 and 2. Independent t-tests compared true negatives to false negatives, which revealed no significant differences in clinicians' coding of anxiety ($p < .268$), overactivity ($p < .056$), or disruptive behaviors ($p < .162$) on Module 1. Significant differences, however, were noted in Module 2 for disruptive behaviors ($p < .01$), with scores higher for false negatives than true negatives, but not for overactivity ($p < .118$) or anxiety ($p < .168$).

Discussion

The BOSA was initially developed by the ADOS-2 authors during the COVID-19 pandemic. At that time, there was a need to observe symptoms of autism through a structured observation because the necessary use of PPE invalidated the ADOS-2. Results from the BOSA validation study indicated that the BOSA had strong sensitivity (0.86-0.98) and specificity (0.70-1.00) in a university-based American sample (Dow et al., 2021). Surprisingly, these numbers approach the psychometric properties produced by the ADOS-2, which is considered the gold-standard diagnostic measure (i.e., sensitivity [0.79-0.98] and specificity [0.69-0.86] depending on the ASD cutoff selected (Lord et al., 2012). To my knowledge, follow-up studies on BOSAs performance, including a replication, have not been conducted. The current study fills this knowledge gap and serves as preliminary evidence to support the BOSA's use as a level-2, direct-observation screening tool for Spanish-speaking individuals at-risk for ASD.

BOSA psychometric properties in Spanish-speaking individuals

In the current study, the BOSA demonstrates strong sensitivity across the ADOS-2 modules except for Module 2. Specifically, the Toddler Module, which is supposed to be used for very young children (e.g., under 30 months), and Module 1, which is used for those who are non-speaking or minimally verbal, demonstrated sufficient sensitivity (i.e., at or above 0.70) to detect those at-risk for ASD. In contrast, the Toddler Module and Module 1 demonstrated specificity under 50%, suggesting significant caution should be used in interpreting the results of these modules as the ability of the cut-off score to differentiate ASD from other disorders is less than chance. While specificity was low, the measure's sensitivity in young children is encouraging if used as a screener. Moreover, the BOSA covers an age range rarely covered by other screeners, with only one other caregiver report measure having proven utility in this range

(i.e., SRS, ages >2;6). Results from the modules often used on more verbally fluent participants found the BOSA detected about 90% of the individuals with fluent speech (e.g., Module 3) who were later diagnosed with ASD, further supporting its possible screening utility. Similar strong sensitivity was found in Module 4, though further clinical conclusions could not be drawn due to the limited sample size. In addition to strong sensitivity, both modules used for older adolescents and adults had strong specificity. BOSAs given to children with phrase speech (i.e., Module 2) produced the lowest sensitivity but strong specificity (0.93), possibly due to the increased number of typically developing controls. In summation, while a high proportion of individuals were identified with ASD using the BOSA (sensitivity), specificity remained variable across modules. Clinicians using this instrument in Spanish-speaking samples should be aware of this limitation, as the overall accuracy is varies from findings in the initial US validation sample (Dow et al., 2021). Furthermore, for those individuals with language skills beyond single words and simple phrases, the BOSA fills a much-needed gap in direct-observation screening tools for autism in Spanish-speaking samples. However, clinicians might consider using other tools for children with phrase speech to assess autism symptomology and risk accurately.

Given the lack of effective level-2 screeners across language levels, the BOSA shows promise as a future screener. However, clinicians should be cautious when using the BOSA in clinically referred samples. The current sample included a high number of individuals who were not diagnosed with any condition in addition to "typical controls." Unfortunately, the category in which I collected data was labeled as "not diagnosed," therefore, I was unable to know the exact number of controls. Future studies should investigate the BOSAs performance in large, referred samples of children and adults at risk for ASD who also present with high levels of non-ASD related disorders (e.g., ADHD, anxiety, Language Disorders, Intellectual and Developmental

Disorder, and Global Developmental Delays). Finally, the BOSA should not be used as a diagnostic tool but rather as a source of information to be used in combination with other direct and indirect assessments to determine risk and diagnostic conclusions.

BOSA as a screener

For diagnostic tools to be considered valid, both sensitivity and specificity should be above 0.70 (Marks et al., 2008). This was not found in this sample. Screeners, however, generally prioritize sensitivity over specificity (Dow et al., 2021). For these reasons, the BOSA could be considered an effective screening tool rather than a diagnostic instrument. As is true for most screeners, the BOSA has limitations that suggest that it should not be used in isolation, especially in children with phrase speech.

Despite these limitations, the BOSA fills many gaps in current ASD screeners. First, it can be administered by someone with minimal training and takes very little time to administer. Though those coding the BOSA should be ADOS-2 trained, exploratory findings from this study revealed clinicians do not need to be research reliable to be relatively accurate when rating the BOSA. Second, the BOSA is unlike most level-2 screeners because it uses direct observational methods. These methods provide structured activities, allow clinicians to view parent-child interactions, and provide valuable insights into their interactions, which adds potentially essential clinical information that other screeners do not provide. Third, the BOSA can be conducted in-home or in clinical settings. However, if conducted at home, logistical and technological barriers might be imposed. For example, BOSA kits must be sent to the family's home and returned to the clinic. This flexibility might alleviate some barriers to access to care, especially in low-income Spanish-speaking countries such as Argentina and Chile. Finally, and probably most clinically relevant, the BOSA encompasses a broad age range (i.e., 12 months through

adulthood). At this point, most direct observation screening tools are limited to very young children or those with minimal speech. In fact, the BOSA demonstrated some specific strengths in those with more fluent language. Based on findings from this study and the English-speaking validation sample (Dow et al., 2021), the BOSA has strong sensitivity for those with expressive language levels equivalent to or above that of a four-year-old child. These findings support the potential clinical usage of the BOSA in both English and Spanish-speaking samples.

BOSA false positives and false negatives

Like every other measure, autism screeners and diagnostic tools are susceptible to false positives and negatives. In the case of most screeners, especially level-2 screeners that focus on a specific disorder, it is ideal to minimize false negatives (i.e., missing children who have ASD). Moreover, sensitivity is often optimized, erring on the side of identifying all those suspected of the disorder. Even though false positives might arise, those that screen positive in most cases would go on to a more comprehensive evaluation. In this Spanish-speaking sample, a larger proportion of false positives were noted compared to false negatives. This might appear problematic as it can lead to increased caregiver stress and concern; however, those screened as false positives were likely to be diagnosed with other disorders (e.g., language disorder, intellectual disability, or global developmental disability). Thus, due to the false positive screen on the BOSA, these children were identified as having a delay, and while it was not ASD, they were directed to appropriate services.

Like Havdahl (2016), the current study determined that an increased number of EBPs was related to a higher likelihood of false positives on the BOSA. There was a greater proportion of false positives in lower modules (i.e., Toddler Module and Module 1) compared to the other modules. Others have found that, in ASD, false positives can also result when the individual has

extremely low intelligence quotients or significant developmental delays. Unfortunately, developmental/cognitive testing was not available for the complete sample, though it is suspected that, like the ADOS-2, individuals with nonverbal mental ages below 12 and 18 months might result in a false positive on the BOSA. Thus, it is essential for clinicians using the BOSA to administer developmental or cognitive assessments to aid in interpreting findings from the BOSA.

Another area that has recently gained attention is understanding false negative results on screening and diagnostic instruments. False negative results are less common than false positives, yet investigations into how they occur might provide insight into instrument refinement that improves accuracy to adequately capture symptomology (Øien et al., 2018; Schjølberg et al., 2021). In this sample, comparisons between true and false negatives did not reveal significant differences related to cognitive profiles, adaptive skills, or total EBPs. However, item-level analyses for false negative children with phrase speech (i.e., Module 2) showed increased disruptive behavior scores compared to children who truly did not have autism. Clinicians noted phenotypic differences in autistic children, but they did not reach the recommended cutoffs associated with risk for ASD on the BOSA. Perhaps the shortened observation period of the BOSA potentially led to diagnostic overshadowing, and clinicians falsely attributed behavioral differences to other symptomology (i.e., disruptive behavior) rather than ASD core symptoms. This finding is consistent with much of the literature on false negative findings, noting that social-communication difficulties might be present in early development but might not be easily detected or accurately attributed to autism symptomology (Øien et al., 2018; Schjølberg et al., 2021). Further work is needed to explore how to improve brief, direct observations that might elicit even more apparent differences in core ASD symptoms, especially

in children with phrase speech.

BOSA Clinically Trained (CT) clinicians vs. Research Reliable (RR) clinicians

Like many screening and diagnostic tools that require a clinician to identify and code specific behaviors, it is important to determine how much training is needed to be accurate. The ADOS-2 has two levels of training: a basic level of training after a clinician is deemed "Clinically Trained" and a more comprehensive training after which the clinician is determined to be "Research- Reliable (RR)." In a research context, studies may use RR clinicians to determine caseness for ASD, ensuring those being studied have ASD. In most clinical contexts, few clinicians reach research reliability. In this study, only a quarter of the clinicians had reached research reliability. Given this percentage, few conclusions could be drawn when distributed across modules related to the level of training needed to rate the BOSA due to the small cell sizes. However, when looking across all ADOS-2 modules, RR clinicians demonstrated significantly higher accuracy overall compared to CT. Further work is needed to understand better the relationship between training level and the impact on accuracy across each module. That being said, findings from this study demonstrated that those only clinically trained on the ADOS-2 fairly accurately captured autism symptomology on the BOSA. Since the BOSA is based on ADOS scoring, those who are CT might consider calibration on coding with RR clinicians to improve reliability. The BOSA allows the administration to be recorded and coded at a later time. This flexibility allows calibration discussions to be held at a more convenient and conducive time for clinicians, which might boost the overall accuracy of the BOSA results.

Limitations

This study has several limitations. First, clinical diagnoses were not independently verified or confirmed and often may have used the BOSA observations for diagnostic

determination. Second, the data were not collected as part of a well-controlled study, which impacts the ability to ensure standardization and fidelity of the BOSA across sites. Clinicians reported their ADOS-2 training and reliability status; however, detailed information related to their training and experiences, such as the duration and quality of their training, amount of experience conducting the ADOS-2, assessing ASD, as well as their ongoing reliability and calibration on the ADOS-2, was not collected. Finally, the BOSA was administered in some cases for diagnostic decision-making and in others as a control for an ongoing validation study. Differences in sample selection, including the proportion of typically developing controls, might impact the psychometric properties of the measure. Despite these limitations, the current study will provide valuable insights into the future clinical use of the BOSA as a level-2 screener in Spanish-speaking samples.

Conclusions

To my knowledge, only one direct observation level-2 screening instrument, the ADEC, has demonstrated utility in Spanish-speaking communities (Alonso-Esteban et al., 2020); however, the instrument was limited to toddler-aged children. The current study provides preliminary evidence for the use of the BOSA as a level-2 screener in Spanish-speaking populations across a broad age range (i.e., toddlers through adulthood). Though this study provides evidence for the use of the BOSA as a screener, future studies that follow formal translation processes and replicate these findings in larger clinically-referred samples are needed.

Similar to other screening and diagnostic measures, increased levels of EBPs impact BOSA results. In many cases, a brief observation might not be sufficient to accurately observe and rate ASD symptoms. Therefore, clinicians might consider using an instrument that collects data over a longer observation period, such as ADOS-2, or collecting BOSA data and combining

other assessment and questionnaire data from multiple sources. Research suggests that combining observational methods with caregiver reports increases diagnostic accuracy (Charman & Gotham, 2013; Kanne et al., 2018; Lord et al., 2022). For this reason, clinicians might consider using the BOSA in combination with traditional screening tools like the M-CHAT R/F, SCQ, or SRS-2 to triage autism symptomology. This potential triage approach allows caregivers the flexibility to complete measures within their home while also providing meaningful direct observations of parent-child interactions rarely collected in the clinical context. In addition, this approach might alleviate the burden of time and travel to specialty diagnostic centers for families, which is especially important in low-income and less-resourced areas. Triaging also might decrease burgeoning waitlists by assessing only those at risk for ASD and triaging less symptomatic patients to more appropriate clinics. Consequently, future research should be conducted to examine if combining autism screening instruments increases accuracy in screening and has any impact on managing diagnostic waitlists. Future endeavors might determine if scoring could be done on items other than ADOS codes and reduce the need for ADOS training (and also purchasing ADOS protocols).

Table 1
Frequency counts of BOSA observations in the sample by the ADOS module.

	Toddler		Module 1		Module 2		Module 3		Module 4	
	ASD	Non-	AS	Non-	ASD	Non-	AS	Non-ASD	ASD	Non-ASD
n	17	30	44	81	13	30	18	39	7	16

Table 2
Sample BOSA participants per ADOS-2 Module.

ADOS- 2 Module	*n*	Males with ASD	Mean age (*SD*)	Age range (in years)
Toddler	47	12	2.15 (0.39)	1.33-3.17
Module 1	126	31	3.74 (1.17)	1.92-9.00
Module 2	43	8	5.38 (2.36)	3.00-14.00
Module 3	57	15	9.35 (3.05)	4.00-20.17
Module 4	22	5	27.38 (11.14)	13.75-49.33

Note: Number of participants with an ASD diagnosis.

Table 3
Demographic information for the sample.

Participant Characteristics	(*n*)	(%)
Male	206	69.80
ASD	99	33.56
Race		
White	250	84.75
South American Indian	23	7.83
Multiracial	14	5.08
Black	1	0.34
Unknown	6	2.03
Non-ASD Diagnoses	196	66.44
Language Disorder	36	18.37
ADHD	5	2.55
IDD/GDD	30	15.31
Anxiety/Depression	2	1.23
No Diagnosis/Undetermined	114	58.16
Other	9	4.59

Note: ADHD=Attention Deficit Hyperactivity Disorder, IDD= Intellectual and Developmental Disorder, GDD=Global Developmental Delay. "Other" diagnoses include various genetic and social-emotional disorders.

Table 4

Adaptive and cognitive assessment data by ADOS-2 module.

	Toddler	Module 1	Module 2	Module 3	Module 4
Developmental Testing (*n)*	8	33	-	-	-
Standard Score *m* (*SD*)	60.75 (10.70)	68.88 (13.63)			
Best Estimate IQ (*n)*	-	10	15	23	-
Standard Score *m* (*SD*)		59.90 (12.92)	82.07 (25.27)	86.65 (20.81)	
Adaptive Skills (*n)*	17	90	20	24	2
Standard Score *m* (*SD*)	72.47 (14.20)	64.23 (12.93)	74.00 (10.70)	73.75 (11.73)	65.00 (2.83)

Note: Developmental testing was conducted using the Bayley Scales of Infant and Toddler Development. Best Estimate Intelligence Quotient (IQ) uses full-scale standard scores from the DAS (Differential Ability Scales), WISC (Wechsler Intelligence Scale for Children), Wechsler Abbreviated Scale of Intelligence (WASI), or WPPSI (Wechsler Preschool and Primary Scale of Intelligence). Adaptive scores consisted of standard scores from the Vineland Scales of Adaptive Behaviors using the Adaptive Behavior Composites (ABC).

Table 5
Clinician profession information.

Clinician Characteristics	(%)
Profession	
Licensed Psychologists	42.00
Child Neurologists	46.80
SLP	11.20

Note: SLP=Speech and Language Pathologists.

Table 6

BOSA Psychometric Properties from the United States validation sample (Dow et al., 2021).

	AUC	Sensitivity (%)	Specificity (%)	Recommended cutoff	Range of concern
Toddler	.96	96	83	6	0-3 Little-to-No 4-5 Mild-to-Moderate 6+ Moderate-to-Severe
Module 1	.97	91	100	5	0-4 Little-to-No 5-8 Mild-to-Moderate 9+ Moderate-to-Severe
Module 2	.87	91	74	9	0-6 Little-to-No 7-8 Mild-to-Moderate 9+ Moderate-to-Severe
Module 3	.91	86	70	6	0-3 Little-to-No 4-5 Mild-to-Moderate 6+ Moderate-to-Severe
Module 4	.98	98	93	3	0-2 Little-to-No 3-4 Mild-to-Moderate 5+ Moderate-to-Severe

Table 7

Psychometric properties of the BOSA by ADOS-2 modules for the Spanish-speaking sample.

	Toddler	Module 1	Module 2	Module 3	Module 4
Sensitivity	0.82	0.70	0.46	0.89	1.00
Specificity	0.47	0.44	0.93	0.79	0.73
PPV	0.47	0.41	0.75	0.67	0.64
NPV	0.82	0.73	0.80	0.94	1.00
Accuracy (*%*)	59.6	54.0	79.1	82.5	81.8

Note: Negative Predictive Value = NPV, and Positive Predictive Value = PPV.

Table 8
Psychometrics by clinician training across ADOS-2 modules.

	Toddler		Module 1		Module 2		Module 3		Module 4	
	RR	CT	RR	CT	RR	CT	RR	CT	RR	CT
n	24	23	27	99	12	35	8	49	4	18
Sensitivity	0.80	0.86	0.79	0.67	*	0.46	1.00	0.85	*	1.00
Specificity	0.79	0.19	0.85	0.38	1.00	0.91	0.67	0.81	1.00	0.64
PPV	0.73	0.32	0.85	0.32	*	0.75	0.83	0.61	*	0.64
NPV	0.85	0.75	0.79	0.72	1.00	0.74	1.00	0.94	1.00	1.00
Accuracy (%)	79.2	39.1	81.5	46.5	33.3	74.3	87.5	81.6	100	77.8

Note: Research Reliable=RR. Clinically Trained=CT. Negative Predictive Value = NPV, and Positive Predictive Value = PPV. *Unable to be calculated due to lack of false negatives.

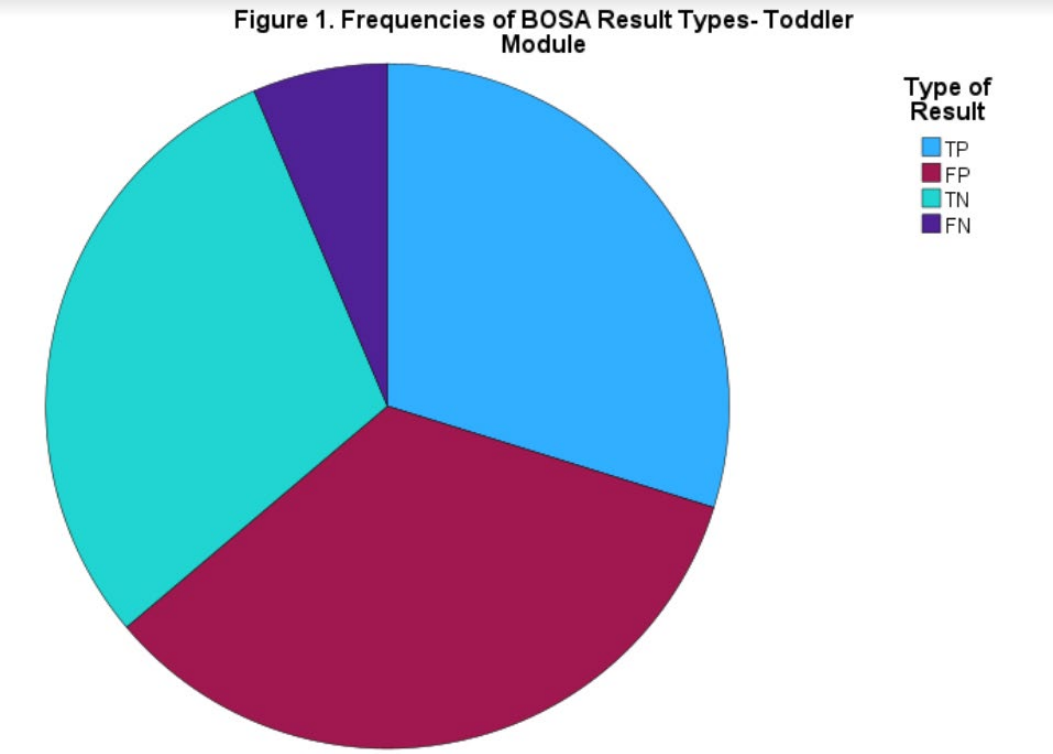
Table 9
Results of the logistical regression for false positives.

	β	*p*-value	Exp(B)
Age	-0.003	0.524	0.997
EBP Total Score*	0.422	0.003	1.525
ADOS-2 Module*	-	0.014	-
Constant	-1.027	0.452	0.358

*p-value is significant, indicating a significant difference between the group.

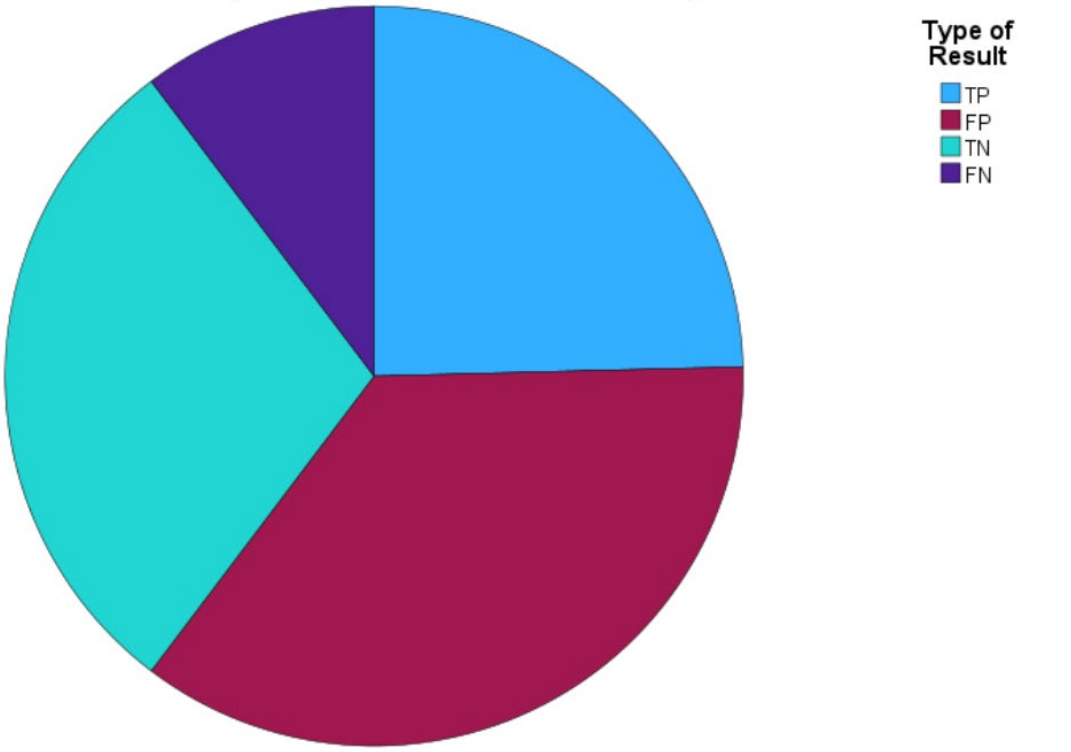
Table 10
Results of the logistical regression for false negatives.

	β	*p*-value	Exp(B)
Age	.004	.680	1.004
EBP Total Score	.286	.155	1.332
ADOS-2 Module	-	.297	-
Constant	-22.866	.998	.000

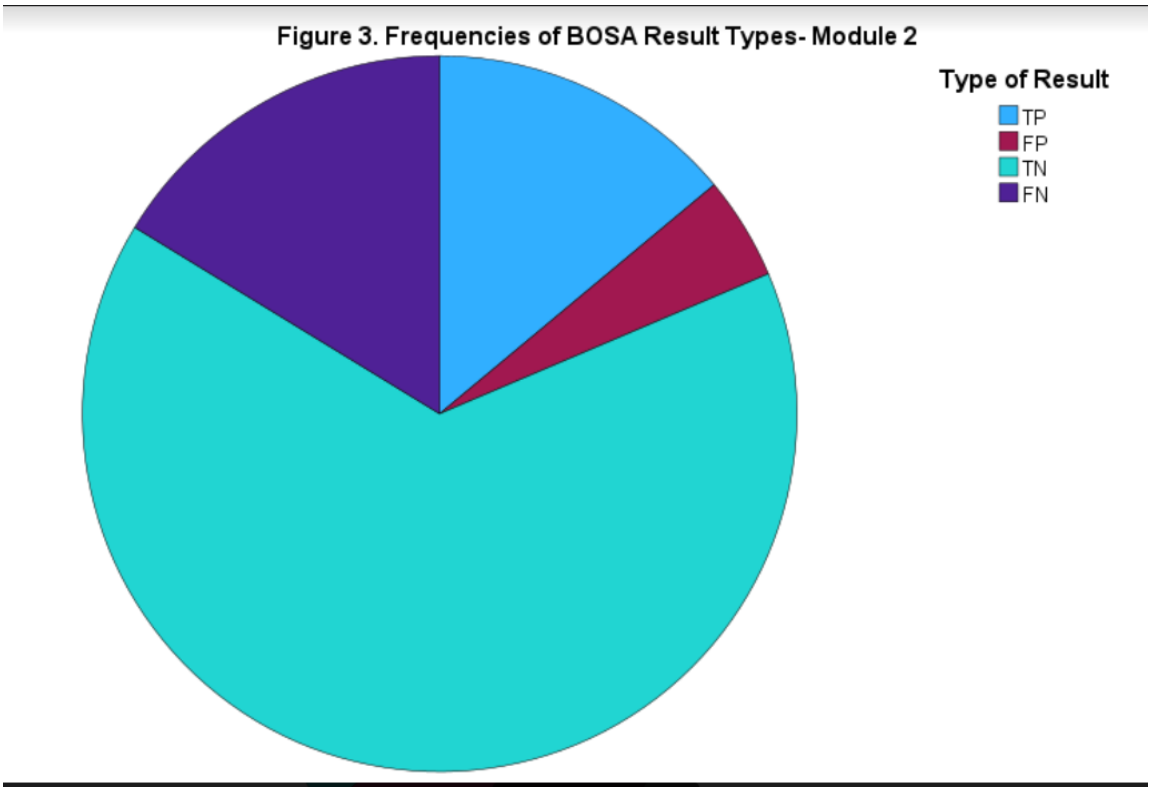


Note: False Positives = FP; True Positives = TP, False Negatives = FN; True Negatives = TN

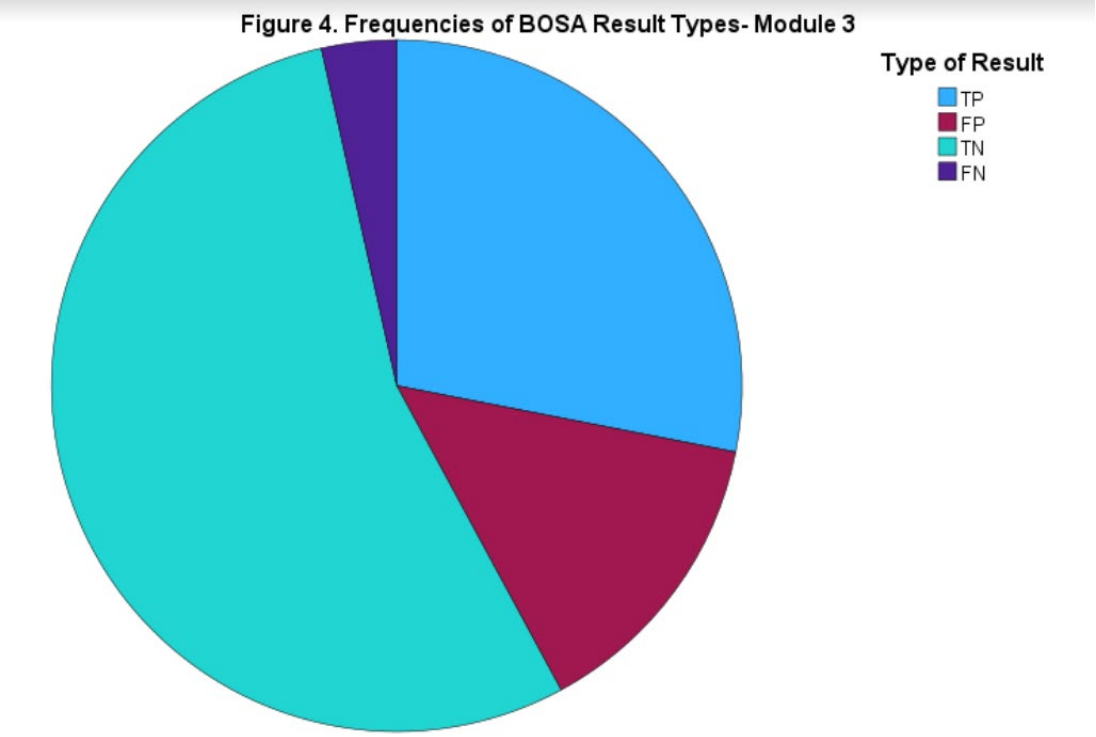
Figure 2. Frequencies of BOSA Result Types- Module 1



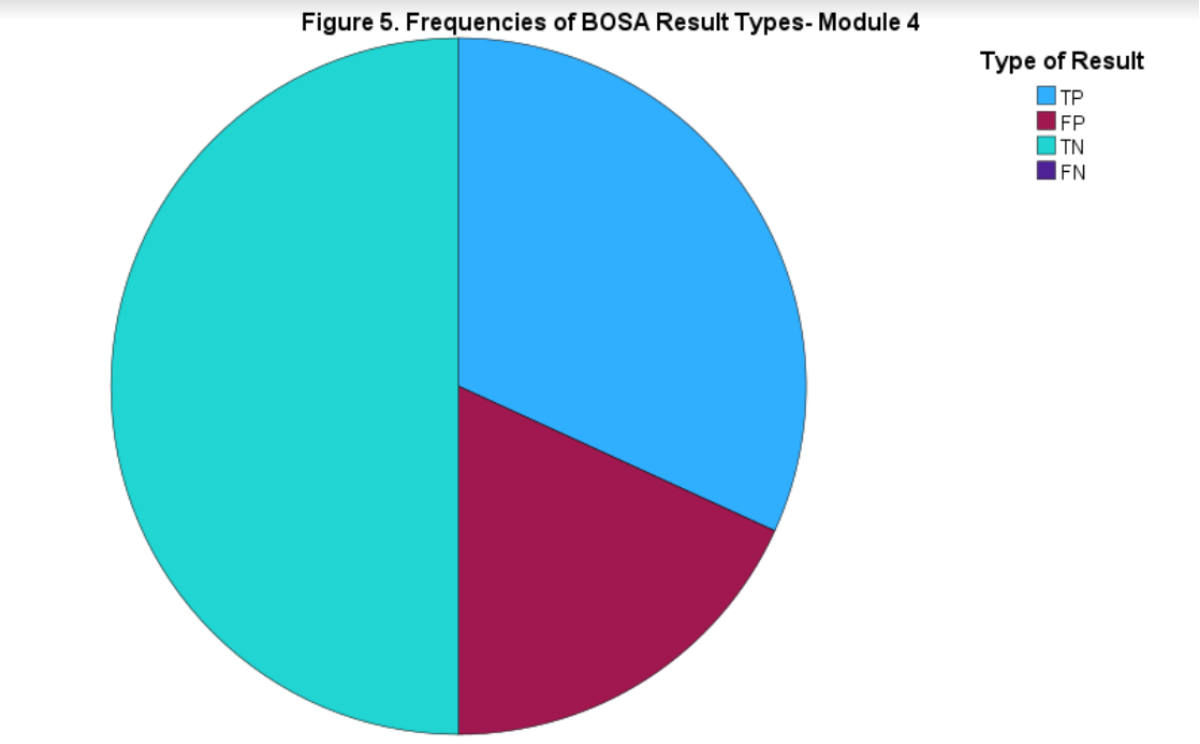
Note: False Positives = FP; True Positives = TP, False Negatives = FN; True Negatives = TN



Note: False Positives = FP; True Positives = TP, False Negatives = FN; True Negatives = TN



Note: False Positives = FP; True Positives = TP, False Negatives = FN; True Negatives = TN



Note: False Positives = FP; True Positives = TP, False Negatives = FN; True Negatives = TN

References

Adiani, D., Schmidt, M., Wade, J., Swanson, A. R., Weitlauf, A., Warren, Z., & Sarkar, N. (2019). Usability Enhancement and Functional Extension of a Digital Tool for Rapid Assessment of Risk for Autism Spectrum Disorders in Toddlers Based on Pilot Test and Interview Data. In M. Antona & C. Stephanidis (Eds.), *Universal Access in Human-Computer Interaction. Multimodality and Assistive Environments* (pp. 13–22). Springer International Publishing. https://doi.org/10.1007/978-3-030-23563-5_2

Aiello, R., Ruble, L., & Esler, A. (2017). National Study of School Psychologists' Use of Evidence-Based Assessment in Autism Spectrum Disorder. *Journal of Applied School Psychology*, *33*(1), 67–88. https://doi.org/10.1080/15377903.2016.1236307

Al Maskari, T. S., Melville, C. A., & Willis, D. S. (2018). Systematic review: Cultural adaptation and feasibility of screening for autism in non-English speaking countries. *International Journal of Mental Health Systems*, *12*(1), 22. https://doi.org/10.1186/s13033-018-0200-8

Alonso-Esteban, Y., Marco, R., Hedley, D., Uljarević, M., Barbaro, J., Canal-Bedia, R., & Alcantud-Marín, F. (2020). Screening instruments for early detection of autism spectrum disorder in Spanish speaking communities. *Psicothema*, *32*(2), 245–252. https://doi.org/10.7334/psicothema2019.340

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

Beacham, C., Reid, M., Bradshaw, J., Lambha, M., Evans, L., Gillespie, S., Klaiman, C., & Richardson, S. S. (2018). Screening for Autism Spectrum Disorder: Profiles of Children Who Are Missed. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *39*(9), 673–682. https://doi.org/10.1097/DBP.0000000000000607

Blacher, J., Cohen, S. R., & Azad, G. (2014). In the eye of the beholder: Reports of autism symptoms by Anglo and Latino mothers. *Research in Autism Spectrum Disorders*, *8*(12), 1648–1656. https://doi.org/10.1016/j.rasd.2014.08.017

Brewer, N., Young, R. L., & Lucas, C. A. (2020). Autism Screening in Early Childhood: Discriminating Autism From Other Developmental Concerns. *Frontiers in Neurology*, *11*. https://www.frontiersin.org/article/10.3389/fneur.2020.594381

Brian, J. A., Zwaigenbaum, L., & Ip, A. (2019). Standards of diagnostic assessment for autism spectrum disorder. *Paediatrics & Child Health*, *24*(7), 444–451. https://doi.org/10.1093/pch/pxz117

Bujang, M. A., & Adnan, T. H. (2016). Requirements for minimum sample size for sensitivity and specificity analysis. *Journal of clinical and diagnostic research: JCDR, 10*(10), YE01.

Camodeca, A. (2022). Diagnostic Utility of the Gilliam Autism Rating Scales-3rd Edition Parent Report in Clinically Referred Children. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05483-5

Cantor-Cutiva, L. C., Jiménez-Chala, E.-A., Bottalico, P., & Hunter, E. J. (2021). Bilingualism and Voice Production. Differences Between Bilingual Latin-American Spanish- English Female Speakers and Monolingual Spanish Female Speakers During Spanish Productions. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2021.04.026

Carruthers, S., Kinnaird, E., Rudra, A., Smith, P., Allison, C., Auyeung, B., Chakrabarti, B., Wakabayashi, A., Baron-Cohen, S., Bakolis, I., & Hoekstra, R. A. (2018). A cross-cultural study of autistic traits across India, Japan and the UK. *Molecular Autism*, *9*, 52. https://doi.org/10.1186/s13229-018-0235-3

Carvajal, D. N., & Rowe, P. C. (2010). Sensitivity, specificity, predictive values, and likelihood ratios. *Pediatrics in Review*, *31*(12), 511–513. https://doi.org/10.1542/pir.31-12-511

Cederberg, C. D., Gann, L. C., Foley-Nicpon, M., & Sussman, Z. (2018). ASD Screening Measures for High-Ability Youth With ASD: Examining the ASSQ and SRS. *Gifted Child Quarterly*, *62*(2), 220–229. https://doi.org/10.1177/0016986217752098

Chang, J., Gilman, S. R., Chiang, A. H., Sanders, S. J., & Vitkup, D. (2015). Genotype to phenotype relationships in autism spectrum disorders. *Nature Neuroscience*, *18*(2), 191–198. https://doi.org/10.1038/nn.3907

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. https://doi.org/10.1111/j.1475-3588.2012.00664.x

Chawner, S. J. R. A., Owen, M. J., Holmans, P., Raymond, F. L., Skuse, D., Hall, J., & van den Bree, M. B. M. (2019). Genotype–phenotype associations in children with copy number variants associated with high neuropsychiatric risk in the UK (IMAGINE-ID): A case-control cohort study. *The Lancet Psychiatry*, *6*(6), 493–505. https://doi.org/10.1016/S2215-0366(19)30123-3

Chesnut, S., Wei, T., Barnard-Brak, L., & Richman, D. (2017). *A meta-analysis of the social communication questionnaire: Screening for autism spectrum disorder*. *21*(8). https://doi.org/10.1177/1362361316660065

Chiarotti, F., & Venerosi, A. (2020). Epidemiology of Autism Spectrum Disorders: A Review of Worldwide Prevalence Estimates Since 2014. *Brain Sciences*, *10*(5), 274. https://doi.org/10.3390/brainsci10050274

Chojnicka, I., & Pisula, E. (2017). Adaptation and Validation of the ADOS-2, Polish Version.

Frontiers in Psychology, *8*, 1916. https://doi.org/10.3389/fpsyg.2017.01916

Christopher, K., Bishop, S., Carpenter, L. A., Warren, Z., & Kanne, S. (2020). The Implications of Parent-Reported Emotional and Behavioral Problems on the Modified Checklist for Autism in Toddlers. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-020-04469-5

Christopher, K., & Lord, C. (n.d.). Best practice assessments for autism spectrum disorders in schools. *Psychology in the Schools*, *n/a*(n/a). https://doi.org/10.1002/pits.22680

Constantino, J. N., Abbacchi, A. M., Saulnier, C., Klaiman, C., Mandell, D. S., Zhang, Y., Hawks, Z., Bates, J., Klin, A., Shattuck, P., Molholm, S., Fitzgerald, R., Roux, A., Lowe, J. K., & Geschwind, D. H. (2020). Timing of the Diagnosis of Autism in African American Children. *Pediatrics*, *146*(3), e20193629. https://doi.org/10.1542/peds.2019-3629

Constantino, J. & Gruber, C. (2012). Social Responsiveness Scale–Second Edition (SRS-2). Torrance, CA: Western Psychological Services.

Constantino, J., & Frazier, T.(2013). Commentary: The observed association between autistic severity measured by the social responsiveness scale (SRS) and general psychopathology--a response to Hus et al.(2013). *J Child Psychol Psychiatry*, *54*, 6, 695-7. doi: 10.1111/jcpp.12064.

Corona, L. L., Weitlauf, A. S., Hine, J., Berman, A., Miceli, A., Nicholson, A., Stone, C., Broderick, N., Francis, S., Juárez, A. P., Vehorn, A., Wagner, L., & Warren, Z. (2021). Parent Perceptions of Caregiver-Mediated Telemedicine Tools for Assessing Autism Risk in Toddlers. *Journal of Autism and Developmental Disorders*, *51*(2), 476–486.

https://doi.org/10.1007/s10803-020-04554-9

Crane, L., Chester, J. W., Goddard, L., Henry, L. A., & Hill, E. (2016). Experiences of autism diagnosis: A survey of over 1000 parents in the United Kingdom. *Autism: The International Journal of Research and Practice*, *20*(2), 153–162. https://doi.org/10.1177/1362361315573636

Daley, T. C., Singhal, N., & Krishnamurthy, V. (2013). Ethical considerations in conducting research on autism spectrum disorders in low and middle income countries. *Journal of Autism and Developmental Disorders*, *43*(9), 2002–2014. https://doi.org/10.1007/s10803-012-1750-2

Dow, D., Day, T. N., Kutta, T. J., Nottke, C., & Wetherby, A. M. (2020). Screening for autism spectrum disorder in a naturalistic home setting using the systematic observation of red flags (SORF) at 18–24 months. *Autism Research*, *13*(1), 122–133. https://doi.org/10.1002/aur.2226

Dow, D., Holbrook, A., Toolan, C., McDonald, N., Sterrett, K., Rosen, N., Kim, S. H., & Lord, C. (2021). The Brief Observation of Symptoms of Autism (BOSA): Development of a New Adapted Assessment Measure for Remote Telehealth Administration Through COVID-19 and Beyond. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-021-05395-w

DuBay, M., Watson, L. R., Méndez, L. I., & Rojevic, C. (2021). Psychometric Comparison of the English and Spanish Western-Hemisphere Versions of the Modified Checklist for Autism in Toddlers-Revised. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *42*(9), 717–725. https://doi.org/10.1097/DBP.0000000000000968

Ehlers, S., Gillberg, C., & Wing, L. (1999). A Screening Questionnaire for Asperger Syndrome

and Other High-Functioning Autism Spectrum Disorders in School Age Children. *Journal of Autism and Developmental Disorders*, *29*(2), 129–141. https://doi.org/10.1023/A:1023040610384

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcín, C., Montiel-Nava, C., Patel, V., Paula, C. S., Wang, C., Yasamy, M. T., & Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research*, *5*(3), 160–179. https://doi.org/10.1002/aur.239

Esler, A., & Ruble, L. (2015) DSM-5 Diagnostic Criteria for Autism Spectrum Disorder with Implications for School Psychologists, *International Journal of School & Educational Psychology, 3*:1, 1-15, DOI: 10.1080/21683603.2014.890148.

Fombonne, E. (2018). Editorial: The rising prevalence of autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *59*(7), 717–720. https://doi.org/10.1111/jcpp.12941

Georgiades, S., Szatmari, P., Duku, E., Zwaigenbaum, L., Bryson, S., Roberts, W., Fombonne, E., Mirenda, P., Smith, I., Vaillancourt, T., Volden, J., Waddell, C., Thompson, A., & Pathways in ASD Study Team. (2011). Phenotypic overlap between core diagnostic features and emotional/behavioral problems in preschool children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*(10), 1321–1329. https://doi.org/10.1007/s10803-010-1158-9

Gergoudis, K., Weinberg, A., Templin, J., Farmer, C., Durkin, A., Weissman, J., Siper, P., Foss-Feig, J., del Pilar Trelles, M., Bernstein, J. A., Buxbaum, J. D., Berry-Kravis, E., Powell, C. M., Sahin, M., Soorya, L., Thurm, A., Kolevzon, A., & Consortium, D. S. (2020). Psychometric Study of the Social Responsiveness Scale in Phelan–McDermid Syndrome.

Autism Research, *13*(8), 1383–1396. https://doi.org/10.1002/aur.2299

Geschwind, D. H., & State, M. W. (2015). Gene hunting in autism spectrum disorder: On the path to precision medicine. *The Lancet. Neurology*, *14*(11), 1109–1120. https://doi.org/10.1016/S1474-4422(15)00044-7

Gilliam J. E. (2014). *Gilliam Autism Rating Scale* (3rd ed.). Austin, TX: Pro-Ed.

Goin-Kochel, R. P., Mackintosh, V. H., & Myers, B. J. (2006). How many doctors does it take to make an autism spectrum diagnosis? *Autism: The International Journal of Research and Practice*, *10*(5), 439–451. https://doi.org/10.1177/1362361306066601

Gray, K. M., Keating, C. M., Taffe, J. R., Brereton, A. V., Einfeld, S. L., Reardon, T. C., & Tonge, B. J. (2014). Adult Outcomes in Autism: Community Inclusion and Living Skills. *Journal of Autism and Developmental Disorders*, *44*(12), 3006–3015. https://doi.org/10.1007/s10803-014-2159-x

Greene, R. K., Vasile, I., Bradbury, K. R., Olsen, A., & Duvall, S. W. (2021). Autism Diagnostic Observation Schedule (ADOS-2) elevations in a clinical sample of children and adolescents who do not have autism: Phenotypic profiles of false positives. *The Clinical Neuropsychologist*, *0*(0), 1–17. https://doi.org/10.1080/13854046.2021.1942220

Guthrie, W., Wallis, K., Bennett, A., Brooks, E., Dudley, J., Gerdes, M., Pandey, J., Levy, S. E., Schultz, R. T., & Miller, J. S. (2019). Accuracy of Autism Screening in a Large Pediatric Network. *Pediatrics*, *144*(4), e20183963. https://doi.org/10.1542/peds.2018-3963

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S. N. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, *95*, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Harrison, A. J., Long, K. A., Tommet, D. C., & Jones, R. N. (2017). Examining the Role of Race, Ethnicity, and Gender on Social and Behavioral Ratings Within the Autism Diagnostic Observation Schedule. *Journal of Autism and Developmental Disorders*, *47*(9), 2770–2782. https://doi.org/10.1007/s10803-017-3176-3

Havdahl, A., & Bishop, S. (2019). Heterogeneity in prevalence of co-occurring psychiatric conditions in autism. *The Lancet. Psychiatry*, *6*(10), 794–795. https://doi.org/10.1016/S2215-0366(19)30326-8

Havdahl, K. A., Hus Bal, V., Huerta, M., Pickles, A., Øyen, A.-S., Stoltenberg, C., Lord, C., & Bishop, S. L. (2016a). Multidimensional influences on autism symptom measures: Implications for use in etiological research. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(12), 1054-1063.e3. https://doi.org/10.1016/J.JAAC.2016.09.490

Havdahl, K. A., Hus Bal, V., Huerta, M., Pickles, A., Øyen, A.-S., Stoltenberg, C., Lord, C., & Bishop, S. L. (2016b). Multidimensional Influences on Autism Symptom Measures: Implications for Use in Etiological Research. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(12), 1054-1063.e3. https://doi.org/10.1016/j.jaac.2016.09.490

Hedley, D., Nevill, R. E., Monroy-Moreno, Y., Fields, N., Wilkins, J., Butter, E., & Mulick, J. A. (2015). Efficacy of the ADEC in Identifying Autism Spectrum Disorder in Clinically

Referred Toddlers in the US. *Journal of Autism and Developmental Disorders*, *45*(8), 2337–2348. https://doi.org/10.1007/s10803-015-2398-5

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. https://doi.org/10.1111/j.1469-7610.2012.02589.x

Jacobs, D., Steyaert, J., Dierickx, K., & Hens, K. (2020). Parents' views and experiences of the autism spectrum disorder diagnosis of their young child: A longitudinal interview study. *European Child & Adolescent Psychiatry*, *29*(8), 1143–1154. https://doi.org/10.1007/s00787-019-01431-4

Kamp-Becker, I., Albertowski, K., Becker, J., Ghahreman, M., Langmann, A., Mingebach, T., Poustka, L., Weber, L., Schmidt, H., Smidt, J., Stehr, T., Roessner, V., Kucharczyk, K., Wolff, N., & Stroth, S. (2018). Diagnostic accuracy of the ADOS and ADOS-2 in clinical practice. *European Child & Adolescent Psychiatry*, *27*(9), 1193–1207. https://doi.org/10.1007/s00787-018-1143-y

Kanne, S. M., Abbacchi, A. M., & Constantino, J. N. (2009). Multi-informant ratings of psychiatric symptom severity in children with autism spectrum disorders: The importance of environmental context. *Journal of Autism and Developmental Disorders*, *39*(6), 856–864. https://doi.org/10.1007/s10803-009-0694-7

Kanne, S. M., & Bishop, S. L. (2021). Editorial Perspective: The autism waitlist crisis and remembering what families need. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *62*(2), 140–142. https://doi.org/10.1111/jcpp.13254

Kanne, S. M., Carpenter, L. A., & Warren, Z. (2018). Screening in toddlers and preschoolers at risk for autism spectrum disorder: Evaluating a novel mobile-health screening tool.

Autism Research, *11*(7), 1038–1049. https://doi.org/10.1002/aur.1959

Kim, S. H., Joseph, R. M., Frazier, J. A., O'Shea, T. M., Chawarska, K., Allred, E. N., Leviton, A., & Kuban, K. K. (2016). Predictive Validity of the Modified Checklist for Autism in Toddlers (M-CHAT) Born Very Preterm. *The Journal of Pediatrics*, *178*, 101-107.e2. https://doi.org/10.1016/j.jpeds.2016.07.052

Kimple, K. S., Bartelt, E. A., Wysocki, K. L., & Steiner, M. J. (2014). Performance of the Modified Checklist for Autism in Toddlers in Spanish-speaking patients. *Clinical Pediatrics*, *53*(7), 632–638. https://doi.org/10.1177/0009922814522346

Lebersfeld, J. B., Swanson, M., Clesi, C. D., & O'Kelley, S. E. (2021). Systematic Review and Meta-Analysis of the Clinical Utility of the ADOS-2 and the ADI-R in Diagnosing Autism Spectrum Disorders in Children. *Journal of Autism and Developmental Disorders*, *51*(11), 4101–4114. https://doi.org/10.1007/s10803-020-04839-z

Lee, K. S., Chung, S. J., Thomas, H. R., Park, J., & Kim, S. H. (2019). Exploring diagnostic validity of the autism diagnostic observation schedule-2 in South Korean toddlers and preschoolers. *Autism Research: Official Journal of the International Society for Autism Research*, *12*(9), 1356–1366. https://doi.org/10.1002/aur.2125

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What Is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(6), 466–474. https://doi.org/10.1016/j.jaac.2017.03.013

Lopez, K., Marroquin, J. M., & Gutierrez, C. (2020). Methods to Decrease Disparities in Age of Autism Diagnosis and Treatment Access among Latinx Children. *Social Work*, *65*(2), 140–148. https://doi.org/10.1093/sw/swaa012

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule: ADOS-2*. Los Angeles, CA: Western Psychological Services.

Lord, C., Charman, T., Havdahl, A., Carbone, P., Anagnostou, E., Boyd, B., Carr, T., de Vries, P. J., Dissanayake, C., Divan, G., Freitag, C. M., Gotelli, M. M., Kasari, C., Knapp, M., Mundy, P., Plank, A., Scahill, L., Servili, C., Shattuck, P., … McCauley, J. B. (2022). The Lancet Commission on the future of care and clinical research in autism. *Lancet (London, England)*, *399*(10321), 271–334. https://doi.org/10.1016/S0140-6736(21)01541-5

Lord, C., Elsabbagh, M., Baird, G., & Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. *Lancet (London, England)*, *392*(10146), 508–520. https://doi.org/10.1016/S0140-6736(18)31129-2

Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of General Psychiatry*, *63*(6), 694–701. https://doi.org/10.1001/archpsyc.63.6.694

Maenner, M. J. (2021). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018. *MMWR. Surveillance Summaries*, *70*. https://doi.org/10.15585/mmwr.ss7011a1

Maenner, M. J., Shaw, K. A., Baio, J., Washington, A., Patrick, M., DiRienzo, 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.-C., 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. *MMWR Surveillance Summaries*, *69*(4), 1–12. https://doi.org/10.15585/mmwr.ss6904a1

Mandell, D. S., Wiggins, L. D., Carpenter, L. A., Daniels, J., DiGuiseppi, C., Durkin, M. S., Giarelli, E., Morrier, M. J., Nicholas, J. S., Pinto-Martin, J. A., Shattuck, P. T., Thomas, K. C., Yeargin-Allsopp, M., & Kirby, R. S. (2009). Racial/Ethnic Disparities in the Identification of Children With Autism Spectrum Disorders. *American Journal of Public Health*, *99*(3), 493–498. https://doi.org/10.2105/AJPH.2007.131243

Marks, K., Glascoe, F., Aylward, G., Shevell, M., Lipkin, P., & Squires, J. (2008). The thorny nature of predictive validity studies on screening tests for developmental-behavioral problems. *Pediatrics*, *122*(4). https://doi.org/10.1542/peds.2007-3142

Masini, E., Loi, E., Vega-Benedetti, A., Carta, M., Doneddu, G., Fadda, R., & Zavattari, P. (2020). An Overview of the Main Genetic, Epigenetic and Environmental Factors Involved in Autism Spectrum Disorder Focusing on Synaptic Activity. *International Journal of Molecular Sciences*, *21*(21). https://doi.org/10.3390/ijms21218290

Mazurek, M. O., Brown, R., Curran, A., & Sohl, K. (2017). ECHO Autism: A New Model for Training Primary Care Providers in Best-Practice Care for Children With Autism. *Clinical Pediatrics*, *56*(3), 247–256. https://doi.org/10.1177/0009922816648288

McPartland, J. C. (2017). Developing Clinically Practicable Biomarkers for Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(9), 2935–2937. https://doi.org/10.1007/s10803-017-3237-7

McPartland, J. C., Bernier, R. A., Jeste, S. S., Dawson, G., Nelson, C. A., Chawarska, K., Earl, R., Faja, S., Johnson, S. P., Sikich, L., Brandt, C. A., Dziura, J. D., Rozenblit, L.,

Hellemann, G., Levin, A. R., Murias, M., Naples, A. J., Platt, M. L., Sabatos-DeVito, M., … Autism Biomarkers Consortium for Clinical Trials. (2020). The Autism Biomarkers Consortium for Clinical Trials (ABC-CT): Scientific Context, Study Design, and Progress Toward Biomarker Qualification. *Frontiers in Integrative Neuroscience*, *14*, 16. https://doi.org/10.3389/fnint.2020.00016

Medda, J. E., Cholemkery, H., & Freitag, C. M. (2019). Sensitivity and Specificity of the ADOS-2 Algorithm in a Large German Sample. *Journal of Autism and Developmental Disorders*, *49*(2), 750–761. https://doi.org/10.1007/s10803-018-3750-3

Molloy, C. A., Murray, D. S., Akers, R., Mitchell, T., & Manning-Courtney, P. (2011). Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. *Autism*, *15*(2), 143–162. https://doi.org/10.1177/1362361310379241

Montgomery, J. M., Newton, B., & Smith, C. (2008). Test Reviews: Gilliam, J. (2006). *Journal of Psychoeducational Assessment*, *26*(4), 395–401. https://doi.org/10.1177/0734282908317116

Moody, E. J., Reyes, N., Ledbetter, C., Wiggins, L., DiGuiseppi, C., Alexander, A., Jackson, S., Lee, L.-C., Levy, S. E., & Rosenberg, S. A. (2017). Screening for Autism with the SRS and SCQ: Variations across Demographic, Developmental and Behavioral Factors in Preschool Children. *Journal of Autism and Developmental Disorders*, *47*(11), 3550–3561. https://doi.org/10.1007/s10803-017-3255-5

Moseley, K. L., Freed, G. L., Bullard, C. M., & Goold, S. D. (2007). Measuring African-American parents' cultural mistrust while in a healthcare setting: A pilot study. *Journal of the National Medical Association*, *99*(1), 15–21.

Nazneen, N., Rozga, A., Smith, C. J., Oberleitner, R., Abowd, G. D., & Arriaga, R. I. (2015). A

Novel System for Supporting Autism Diagnosis Using Home Videos: Iterative Development and Evaluation of System Design. *JMIR MHealth and UHealth*, *3*(2), e68. https://doi.org/10.2196/mhealth.4393

Norris, M., & Lecavalier, L. (2010). Screening accuracy of Level 2 autism spectrum disorder rating scales. A review of selected instruments. *Autism: The International Journal of Research and Practice*, *14*(4), 263–284. https://doi.org/10.1177/1362361309348071

Øien, R. A., Vivanti, G., & Robins, D. L. (2021). Editorial S.I: Early Identification in Autism Spectrum Disorders: The Present and Future, and Advances in Early Identification. *Journal of Autism and Developmental Disorders*, *51*(3), 763–768. https://doi.org/10.1007/s10803-020-04860-2

Øien, R., Schjølberg, S., Volkmar, F., Shic, F., Cicchetti, D., Nordahl-Hansen, A., Stenberg, N., Hornig, M., Havdahl, A., Øyen, A., Ventola, P., Susser, E., Eisemann, M., & Chawarska, K. (2018). Clinical Features of Children With Autism Who Passed 18-Month Screening. *Pediatrics, 141*(6), doi: 10.1542/peds.2017-3596.

Overton, T., Fielding, C., & de Alba, R. G. (2008). Brief Report: Exploratory analysis of the ADOS revised algorithm: specificity and predictive value with Hispanic children referred for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(6), 1166–1169. https://doi.org/10.1007/s10803-007-0488-8

Pacífico, M. C., de Paula, C. S., Namur, V. S., Lowenthal, R., Bosa, C. A., & Teixeira, M. C. T. V. (2019). Preliminary evidence of the validity process of the Autism Diagnostic Observation Schedule (ADOS): Translation, cross-cultural adaptation and semantic equivalence of the Brazilian Portuguese version. *Trends in Psychiatry and Psychotherapy*, *41*(3), 218–226. https://doi.org/10.1590/2237-6089-2018-0063

Petruccelli, M., Ramella, L., Schaefer, A. J., Sheldrick, R. C., Carter, A. S., Eisenhower, A., Broder-Fingert, S., & Mackie, T. I. (2022). A Taxonomy of Reported Harms in Pediatric Autism Spectrum Disorder Screening: Provider and Parent Perspectives. *Journal of Autism and Developmental Disorders*, *52*(2), 647–673. https://doi.org/10.1007/s10803-021-04964-3

Pierce, K., Gazestani, V. H., Bacon, E., Barnes, C. C., Cha, D., Nalabolu, S., Lopez, L., Moore, A., Pence-Stophaeros, S., & Courchesne, E. (2019). Evaluation of the Diagnostic Stability of the Early Autism Spectrum Disorder Phenotype in the General Population Starting at 12 Months. *JAMA Pediatrics*, *173*(6), 578–587. https://doi.org/10.1001/jamapediatrics.2019.0624

Posserud, M., Lundervold, A.,Gillberg, C. (2009). Validation of the autism spectrum screening questionnaire in a total population sample. *Journal of Autism and Developmental Disorders, 39,* 126-134.

Ramaswami, G., & Geschwind, D. H. (2018). Genetics of autism spectrum disorder. *Handbook of Clinical Neurology*, *147*, 321–329. https://doi.org/10.1016/B978-0-444-63233-3.00021-X

Rea, K. E., Armstrong-Brine, M., Ramirez, L., & Stancin, T. (2019). Ethnic Disparities in Autism Spectrum Disorder Screening and Referral: Implications for Pediatric Practice. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *40*(7), 493–500. https://doi.org/10.1097/DBP.0000000000000691

Robins, D. L., Casagrande, K., Barton, M., Chen, C.-M. A., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the Modified Checklist for Autism in Toddlers, Revised With Follow-up (M-CHAT-R/F). *Pediatrics*, *133*(1), 37. https://doi.org/10.1542/peds.2013-

1813

Rosen, N. E., Lord, C., & Volkmar, F. R. (2021). The Diagnosis of Autism: From Kanner to DSM-III to DSM-5 and Beyond. *Journal of Autism and Developmental Disorders*, *51*(12), 4253–4270. https://doi.org/10.1007/s10803-021-04904-1

Rutter, M., LeCouteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised (ADI-R)*. Los Angeles, CA. Western Psychological Services.

Rylaarsdam, L., & Guemez-Gamboa, A. (2019). Genetic Causes and Modifiers of Autism Spectrum Disorder. *Frontiers in Cellular Neuroscience*, *13*. https://www.frontiersin.org/article/10.3389/fncel.2019.00385

Samadi, S. A., Noori, H., Abdullah, A., Ahmed, L., Abdalla, B., Biçak, C. A., & McConkey, R. (2022). The Psychometric Properties of the Gilliam Autism Rating Scale (GARS-3) with Kurdish Samples of Children with Developmental Disabilities. *Children (Basel, Switzerland)*, *9*(3), 434. https://doi.org/10.3390/children9030434

Sanchez, M. J., & Constantino, J. N. (2020). Expediting clinician assessment in the diagnosis of autism spectrum disorder. *Developmental Medicine and Child Neurology*, *62*(7), 806–812. https://doi.org/10.1111/dmcn.14530

Schopler, E., Van Bourgondien, M., Wellman, G., & Love, S. (2010). Childhood autism rating scale-second edition (CARS2): Manual. Los Angeles: Western Psychological Services.

Schjølberg, S., Shic, F., Volkmar, F., Nordahl-Hansen, A., Stenberg, N., Torske, T., Larsen, K., Riley, K., Sukhodolsky, D., Leckman, J., Chawarska, K., & Øien, R. (2021). What are we optimizing for in autism screening? Examination of algorithmic changes in the M-CHAT. *Autism Research, 15*(2), 296-304. doi: 10.1002/aur.2643. Epub 2021 Nov 26. PMID: 34837355; PMCID: PMC8821132.

Shreffler, J., & Huecker, M. (2022). Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios. *StatPearls*. https://www.statpearls.com/ArticleLibrary/viewarticle/96435

Shulman, C., Esler, A., Morrier, M. J., & Rice, C. E. (2020). Diagnosis of Autism Spectrum Disorder Across the Lifespan. *Child and Adolescent Psychiatric Clinics of North America*, *29*(2), 253–273. https://doi.org/10.1016/j.chc.2020.01.001

Smith, C. J., Rozga, A., Matthews, N., Oberleitner, R., Nazneen, N., & Abowd, G. (2017). Investigating the accuracy of a novel telehealth diagnostic approach for autism spectrum disorder. *Psychological Assessment*, *29*(3), 245–252. https://doi.org/10.1037/pas0000317

Smith, L., Malcolm-Smith, S., & de Vries, P. J. (2017). Translation and cultural appropriateness of the Autism Diagnostic Observation Schedule-2 in Afrikaans. *Autism: The International Journal of Research and Practice*, *21*(5), 552–563. https://doi.org/10.1177/1362361316648469

Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). Vineland adaptive behavior scales (3rd ed.). *Pearson.*

Squires, J., Twombly, E., Bricker, D., & Potter, L. (2009). The ASQ-3 user's guide- 3rd ed., Baltimore, MD. *Brookes.*

Stevanovic, D., Costanzo, F., Fucà, E., Valeri, G., Vicari, S., Robins, D. L., Samms-Vaughan, M., Ozek Erkuran, H., Yaylaci, F., Deshpande, S. N., Deshmukh, V., Arora, N. K., Albores-Gallo, L., García-López, C., Gatica-Bahamonde, G., Gabunia, M., Zirakashvili, M., Machado, F. P., Radan, M., … Knez, R. (2021). Measurement invariance of the Childhood Autism Rating Scale (CARS) across six countries. *Autism Research: Official Journal of the International Society for Autism Research*, *14*(12), 2544–2554.

https://doi.org/10.1002/aur.2586

Stewart, L. A., & Lee, L.-C. (2017). Screening for autism spectrum disorder in low- and middle-income countries: A systematic review. *Autism: The International Journal of Research and Practice*, *21*(5), 527–539. https://doi.org/10.1177/1362361316677025

Thapar, A., & Rutter, M. (2021). Genetic Advances in Autism. *Journal of Autism and Developmental Disorders*, *51*(12), 4321–4332. https://doi.org/10.1007/s10803-020-04685-z

Vorstman, J. A. S., Parr, J. R., Moreno-De-Luca, D., Anney, R. J. L., Nurnberger, J. I., & Hallmayer, J. F. (2017). Autism genetics: Opportunities and challenges for clinical translation. *Nature Reviews. Genetics*, *18*(6), 362–376. https://doi.org/10.1038/nrg.2017.4

Young, R. L. (2007). Autism detection in early childhood: ADEC. Camberwell, VIC: *Australian Council of Educational Researc*h.

Zablotsky, B., Black, L. I., Maenner, M. J., Schieve, L. A., & Blumberg, S. J. (2015). Estimated Prevalence of Autism and Other Developmental Disabilities Following Questionnaire Changes in the 2014 National Health Interview Survey. *National Health Statistics Reports*, *87*, 1–20.

Zander, E., Willfors, C., Berggren, S., Choque-Olsson, N., Coco, C., Elmund, A., Moretti, Å. H., Holm, A., Jifält, I., Kosieradzki, R., Linder, J., Nordin, V., Olafsdottir, K., Poltrago, L., & Bölte, S. (2016). The objectivity of the Autism Diagnostic Observation Schedule (ADOS) in naturalistic clinical settings. *European Child & Adolescent Psychiatry*, *25*(7), 769–780. https://doi.org/10.1007/s00787-015-0793-2

Zuckerman, K. E., Lindly, O. J., & Sinche, B. K. (2015). Parental Concerns, Provider Response, and Timeliness of Autism Spectrum Disorder Diagnosis. *The Journal of Pediatrics*,

166(6), 1431-1439.e1. https://doi.org/10.1016/j.jpeds.2015.03.007

Zwaigenbaum, L., Bauman, M. L., Choueiri, R., Kasari, C., Carter, A., Granpeesheh, D., Mailloux, Z., Smith Roley, S., Wagner, S., Fein, D., Pierce, K., Buie, T., Davis, P. A., Newschaffer, C., Robins, D., Wetherby, A., Stone, W. L., Yirmiya, N., Estes, A., … Natowicz, M. R. (2015). Early Intervention for Children With Autism Spectrum Disorder Under 3 Years of Age: Recommendations for Practice and Research. *Pediatrics*, *136 Suppl 1*, S60-81. https://doi.org/10.1542/peds.2014-3667E

Zwaigenbaum, L., Bauman, M. L., Stone, W. L., Yirmiya, N., Estes, A., Hansen, R. L., McPartland, J. C., Natowicz, M. R., Choueiri, R., Fein, D., Kasari, C., Pierce, K., Buie, T., Carter, A., Davis, P. A., Granpeesheh, D., Mailloux, Z., Newschaffer, C., Robins, D., … Wetherby, A. (2015). Early Identification of Autism Spectrum Disorder: Recommendations for Practice and Research. *Pediatrics*, *136 Suppl 1*, S10-40. https://doi.org/10.1542/peds.2014-3667C

Zwaigenbaum, L., & Warren, Z. (2021). Commentary: Embracing innovation is necessary to improve assessment and care for individuals with ASD: a reflection on Kanne and Bishop (2020). *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *62*(2), 143–145. https://doi.org/10.1111/jcpp.13271