Developing Autism Screening Criteria for the Brief Infant Toddler Social Emotional Assessment (BITSEA)

Ivy Giserman Kiss, M.A.[1] **[Doctoral Candidate]**, **Melanie S. Feldman, B.A.**[1] **[Doctoral Student]**, **R. Christopher Sheldrick, Ph.D.**[2] **[faculty member]**, and **Alice S. Carter, Ph.D.**[1] **[faculty member]**

[1]Department of Psychology, University of Massachusetts Boston, Boston, MA

[2]Developmental-Behavioral Pediatrics, Tufts University School of Medicine, Boston, MA

Abstract

There is a critical need for evidence-based, broadband behavioral, and ASD screening measures for use in pediatric and early educational settings to ensure that young children at risk for developing social-emotional disorders and/or ASD are provided with early intervention services to optimize long-term outcomes. The BITSEA is a 42-item screener designed to identify social-emotional/behavioral problems and delays/deficits in social-emotional competence among 11-to-48-month-olds; 19 items describe behaviors consistent with ASD. Secondary data analysis was employed to develop cut-scores for ASD subscales using Receiver Operating Curves, discriminating children with (n=223) and without (n=289) ASD. Cut-scores demonstrated moderate-to-high discriminative power, sensitivity, specificity, and PPV. Findings highlight feasibility of using a broadband social-emotional competence and behavior problem screener to improve early detection of ASD.

Keywords

autism spectrum disorder; autism; screening; BITSEA; early detection

The benefits of early intervention (EI) services for children diagnosed with ASD (Pickles et al. 2016; Seida et al., 2009; Woods & Wetherby, 2003) coupled with recommendations of the American Academy of Pediatrics (AAP) and Centers for Disease Control (CDC) for frequent developmental surveillance and screening (AAP, 2006; Baio, 2012), demonstrates a clear demand for effective screening tools designed to detect the early-emerging symptoms of autism spectrum disorder (ASD). Extensive research over the past several decades has identified numerous early symptoms of the disorder, which has laid the foundation for existing broadband screening measures as well as autism-specific screening measures (Jones, Gliga, Bedford, Charman, & Johnson, 2014; Ozonoff et al., 2011; Young, Brewer, &

Correspondence concerning this article should be addressed to Alice S. Carter, Ph.D. at alices.carter@umb.edu, Phone: 617-287-6375, or 100 Morrissey Blvd., Boston, MA, 02125. Fax: 617-287-6336.
There have been no changes in the author affiliation since the time of the study.

Conflicts of Interest: The fourth author receives royalties related to the licensing of the BITSEA. The first three authors declare that they have no conflict of interest.

Pattison, 2003). However, despite many children presenting ASD symptoms between the ages of 12 and 18 months (see review: Barbaro & Dissanayake, 2009), most children do not receive a diagnosis until they are three or four years old, and children from economically disadvantaged or racial/ethnic minority groups are diagnosed even later, on average (Al-Qabandi, Gorter, & Rosenbaum, 2011; Mandell, Listerud, Levy, & Pinto-Martin, 2002). Therefore, there is a need for a measure that can screen broadly for social-emotional and behavioral disorders, in line with the AAP guidelines to screen early and often (AAP, 2006), while simultaneously screening for ASD, with the goal of identifying children at younger ages.

The Brief Infant-Toddler Social and Emotional Assessment (BITSEA) (Briggs-Gowan & Carter, 2006), is a 42-item screener developed to identify social-emotional and behavior problems and delays in the acquisition of competencies between 11- and 48 months of age. The BITSEA has been shown to have excellent psychometric properties in original (Briggs-Gowan, Carter, Irwin, Wachtel, & Cicchetti, 2004) and replication studies (e.g., Kruizinga et al., 2012), strong prediction of concurrent psychiatric disorders (Briggs-Gowan et al., 2013), and good prediction to parent- and teacher- reported school-aged psychopathology (Briggs-Gowan & Carter, 2008).

The BITSEA contains 19 items that are consistent with early-emerging ASD symptomatology (See Table 1). An important feature of the BITSEA is its inclusion of both positively and negatively worded items, relating to both delayed competencies and problem behaviors of ASD. While the BITSEA items that relate to ASD have been recommended to inform clinical judgment, developing cut-scores for the subset of 10 ASD-related problem and 9 ASD-related competence items will allow clinicians to efficiently identify young children who are showing early symptoms of ASD (and subsequently conduct further assessment or make appropriate referrals), while simultaneously screening for early emerging non-ASD social-emotional and behavior problems. Moreover, by eliminating the need for separate social-emotional/behavioral and ASD screeners, the use of a single broadband screener to screen for both ASD and social-emotional/behavioral problems may be more acceptable to pediatric providers and early educators for at least two reasons. First, a single screener increases efficiency. Second, pediatric providers and early educators may feel uncomfortable introducing an ASD-specific screener to families who have not raised any concerns about their children's social-emotional development; thus, inclusion in a broadband tool allows for ASD screening without raising undue concern.

Prospective studies of younger siblings of children diagnosed with ASD (Landa & Garrett-Mayer, 2006; Frohna, 2007; Ozonoff et al., 2010; Yirmiya et al., 2006, Zwaigenbaum et al., 2009) have found that by 12 months of age, children who are later diagnosed with ASD are more likely than those who do not meet diagnostic criteria to exhibit a myriad of symptoms. For instance, infants later diagnosed with ASD are more likely to exhibit abnormal social communication skills such as limited or unusual eye contact (Zwaigenbaum et al., 2005), failure to orient to name call (Frohna, 2007), a lack of use of gestures (Yirmiya et al., 2006), language delays (Landa & Garrett-Mayer, 2006), and visual abnormalities (Bryson et al., 2007; Zwaigenbaum et al., 2005). These studies provide increased evidence of the need for standardized developmental surveillance and screening of infants starting around one year of

age, which is the earliest age at which the risk for ASD can be detected with questionnaires (Reznick, Baranek, Reavis, Watson, & Crais, 2007). A measure such as the BITSEA, which assesses for both social-communicative and behavioral symptoms related to ASD by parent report, is well suited to capture early-emerging symptoms of the disorder (See Table 1).

As public awareness of the prevalence of ASD and knowledge of early-emerging symptoms has grown, so have screening initiatives for young children. The AAP and the CDC recommend that every well-child visit include developmental surveillance, and that healthcare providers routinely administer developmental screeners at the 9-, 18-, and 24- or 30-month visits (AAP, 2006; Baio, 2012). Over the past the past 25 years, a number of ASD-specific screening measures have been developed for use with toddlers, including the Checklist for Autism in Toddlers (CHAT) (Baron-Cohen, Allen, & Gillberg, 1992), the Quantitative Checklist for Autism in Toddlers (Q-CHAT) (Allison et al., 2008), the Modified Checklist for Autism in Toddlers (M-CHAT) (Robins, Fein, Barton, & Green, 2001), the Pervasive Developmental Disorders Screening Test II (PDDST-II) (Siegel, 2004), the Early Screening of Autistic Traits Questionnaire (ESAT) (Dietz, Swinkels, van Daalen, van Engeland, & Buitelaar, 2006), the First Year Inventory (FYI) (Reznick et al., 2007), the Infant Toddler Checklist (ITC) (Wetherby et al., 2004; Wetherby, Brosnan-Maddox, Peace, & Newton, 2008), the Developmental Behavior Checklist – Early Screen (DBC-ES) (Gray & Tonge, 2005), and the Baby and Infant Screen for Children with aUtIsm Traits (BISCUIT) (Matson, Boisjoli, & Wilkins, 2007). While some of these screeners have demonstrated effectiveness in identifying young children at risk for ASD, they vary significantly in accessibility, ease of use, and length of time required to complete. Unlike the above-described measures, the BITSEA (Briggs-Gowan & Carter, 2006) is a brief measure that takes parents approximately five to seven minutes to complete, is written at a fourth to sixth grade reading level, has been internationally validated (e.g., Briggs-Gowan et al., 2004, Kruizinga et al., 2012),) and simultaneously assesses for ASD-specific behaviors as well as more general externalizing, internalizing, and dysregulation problems.

Screening tools can be discussed and compared in relation to four critical psychometric properties, including *sensitivity*, or the proportion of children with ASD who are correctly identified, and *specificity*, or the proportion of children without ASD correctly identified. Although the AAP (2006) recommends that both the sensitivity and specificity of a screening measure be higher than 70%, others argue that the "cut scores" of a screening tool (and thus the relative balance between sensitivity and specificity) should depend on the relative "costs" and "benefits" (i.e., financial, emotional, psychological) of available options for further assessment and intervention (Barton, Dumont-Mathieu, & Fein, 2012; Sheldrick & Garfinkel, in press; Sheldrick, Merchant, & Perrin, 2011; Swets, Dawes, & Monahan, 2000). Given sensitivity, specificity, and prevalence, one can calculate a third critical psychometric property of screening tools, i.e., the *positive predictive value* (PPV), or the proportion of children who meet criteria for a diagnosis out of all of those who screen positive. The overarching goal of this study is to evaluate the feasibility of using the ASD-related items to screen for ASD by evaluating its preliminary sensitivity, specificity, PPV and NPV in a convenience sample.

Methods

Participants

The sample includes participants from three separate projects, two of which only included toddlers diagnosed with ASD using gold-standard measures and techniques (Studies 1 and 2; Blinded for review) and a third that included young children with typical development, developmental delay, and diagnoses of psychopathology other than ASD (Study 3; Blinded for review). See Table 2 for participant sociodemographic information and required diagnostic criteria across the three projects. Participants were recruited through early intervention programs, pediatric practices, infant-parent mental health clinics, and ASD-related networks. Once deemed eligible for the various studies, structured interviews and age-appropriate developmental, language, and social-emotional measures were used to collect data. The overall sample consists of 512 young children (223 in the ASD group, and 289 in the non-ASD group), ranging in age from 15-48 months.

Measures

Brief Infant-Toddler Social Emotional Assessment—The BITSEA is a 42-item brief screening tool aimed to efficiently identify young children who may have social-emotional and behavioral problems and/or delays, or deficits in social-emotional competence. The BITSEA has been used in diverse samples, can be completed in approximately five to seven minutes, and is written at a fourth- to sixth-grade reading level (Briggs-Gowan et al., 2004). There are 19 items on the measure that the authors have identified as consistent with behaviors seen in ASD that fall onto two subscales: the ASD-Problems subscale and the ASD-Competence subscale. The ASD-Problems and ASD-Competence subscales are calculated and prorated if fewer than three items are missing from either scale. In the current sample, the 10 items of the ASD-Problems subscale and the nine items of the ASD-Competence subscale demonstrated adequate reliability (Cronbach's Alpha = .76 and .84, respectively). The ASD-Total subscale score is equal to the value of the ASD-Problem subscale minus the ASD competence.

Mullen Scales of Early Learning—(MSEL; Mullen, 1995) The Mullen is a developmental assessment designed for use with infants and young children from birth to 68 months. The MSEL consists of five different scales: Visual Reception, Fine Motor, Gross Motor, Expressive Language, and Receptive Language, and also generates a composite score, the Early Learning Composite (ELC), that indicates a child's overall developmental level. The measure is norm-referenced and standardized, and has been shown to be an effective psychological assessment for young children with ASD (Akshoomoff, 2006). MSEL scores were evaluated to determine characteristics of children who were correctly identified relative to those who were false negatives and those who were false positives on the BITSEA ASD subscales. The majority of the sample (80%) completed the four subtests (Visual Reception, Fine Motor, Receptive Language, Expressive Language) that comprise the ELC; multiple imputation was used to calculate the 3% of missing MSEL data (using ASD status, age, sex, and available MSEL scales as predictors).

Autism Diagnostic Observation Schedule—(ADOS; Lord et al., 2000). The ADOS is a semi-structured interactive observational tool intended to assess social and communication behaviors of individuals suspected of having an ASD diagnosis. Participants in these samples were administered Modules 1 or 2 based on language ability by research reliable administrators. The tool consists of a number of standardized social presses that are designed to elicit specific behaviors. The ADOS was administered only to the children in the ASD group.

Results

Statistical Analyses

The associations between ASD-Problem, ASD-Competence, and the ASD-Total subscale scores and diagnoses of ASD were described with three contingency tables. Receiver-operating characteristic (ROC) plots were used to determine optimal cut-scores on the BITSEA ASD subscales. ROC plots graph the true positive rate (sensitivity) and false positive rate (1-specificity) of each possible cut-score or threshold. The derived curve formed by the points provides a visual aid in determining the optimal cut-scores according to Youden's Index (e.g., equivalent to the sum of sensitivity and specificity). The cut-scores represent the maximum vertical distance between the ROC curve and the "chance" line (where both the sensitivity and specificity rates are 50%). In other words, maximizing Youden's index specifies a cut-score that detects the given outcome with a degree of accuracy that is farthest from chance (Akobeng, 2007).). The area under the curve (AUC) was calculated to determine the overall discriminative power of the BITSEA in distinguishing children diagnosed with ASD from those who are not. Greater AUC indicates more discriminative power to differentiate children with ASD from those without. By convention, an AUC greater than 0.90 indicates high accuracy for the continuous score, 0.70-0.90 indicates moderate accuracy, 0.50-0.70 low accuracy, and 0.50 chance level accuracy (Swets, 1988).

In addition, we assessed the influence of sampling variation on Youden's index using bootstrapping methods. Bootstrapping is frequently used for calculating confidence intervals of unknown parameters (Zhou & Qin, 2012) and recent research indicates that using bootstrap methods are superior to other methods of assessing sampling variation for Youden's index (Fluss, Faraggi, & Reiser, 2005 Youden, 1950). In the current paper, bootstrapping methods are used to estimate the stability of the cut-scores. Due to the non-representative nature of our data, other criteria for obtaining an optimal cut-score, such as efficiency, misclassification-cost, odds ratio and the kappa index (Fluss et al., 2005) are not considered further in the current paper.

Further, it was hypothesized that children who falsely screened positive for ASD on the BITSEA (false positives) would display significantly lower functioning than those children who were correctly identified as not having the disorder (true negatives). Specifically, children with significant expressive and receptive language delays may use compensatory communication strategies that can present similarly to ASD symptomology (e.g., using another person's hand as a tool). In addition, it was hypothesized that children who falsely screened negative for ASD on the BITSEA would have relatively higher cognitive

functioning than those children who were correctly identified as having the disorder. Specifically, if young children had average to strong language skills, ASD symptoms might not be as clear to a parent and therefore an insufficient number of BITSEA symptoms may be reported for the child to screen positive for ASD.

Findings

ROC curves describing the relationship between BITSEA ASD subscale scores and ASD diagnoses are depicted in Figure 1. The AUC of the ASD-Problem subscale was 0.81, indicating moderate subscale accuracy, and the AUC for the ASD-Competence subscale was 0.92, indicating high subscale accuracy. On the ASD-Problems scale, higher scores indicated more problems. A cut score of 4.5 or greater was identified by Youden's index, displaying a sensitivity of 76% and a specificity of 72%. The cut-score of 4.5 was consistent across 682 of bootstrap 1000 samples, indicating that the cut-score was moderately stable. The 31.8% of bootstrap samples that did not support this cut-score indicate that there is a small probability that thresholds of 3.5, 5.5, or 6.5 may be optimal according to Youden's index. On the ASD competence scale, higher scores indicated greater competence. A cut score of 11.5 or less was identified by Youden's index, displaying a sensitivity of 91% and a specificity of 80%. The cut-score of 11.5 was consistent across 937 of 1000 bootstrap samples, indicating that the cut-score was highly stable. The 6.3% of bootstrap samples that did not support this cut-score indicate that there is a small probability that thresholds of 9.5, 10.5 or 12.5 may be optimal according to Youden's index.

To determine whether a single cut-score comprised of all 19 ASD-related items (ASD-Total subscale) would perform better than either the ASD-Problem or ASD-Competence scales, an ROC plot was calculated to identify the optimal cut-score, sensitivity, specificity, AUC and Youden's Index. The AUC of the ASD-Total subscale was 0.92, indicating high subscale accuracy. Additionally, the Youden's Index indicated that the optimal cut-score for the ASD-Total subscale was greater than or equal to -7.5, which displayed a sensitivity of 89% and a specificity of 79%. The cut-score of -7.5 was consistent across 845 of 1000 bootstrap samples, indicating that the cut-score was stable. The 15.5% of bootstrap samples that did not support this cut-score indicate that there is a small probability that thresholds of -9.5, -8.5, -6.5, -5.5, or -3.5 may be optimal according to Youden's index. Based on these results, the ASD-Competence subscale demonstrated the highest rates of sensitivity and specificity overall. See Table 3 for a comparison of cut-scores across BITSEA ASD subscales. Of note, the ASD-Competence cut score worked equally well across all ages in this convenience sample. Given interest in early detection, we re-ran these analyses for the children who were 24 months and younger. Sensitivity and specificity for all three subscale cut-scores in this subgroup were comparable or stronger (see Table 3).

To explore the clinical characteristics of the children who screened positive despite not receiving an ASD diagnosis (hereafter referred to as false positives) and those who screened negative despite receiving a diagnosis (hereafter referred to as false negatives), the ASD-Competence subscale with a cut-score of less than or equal to 11.5 was applied to the sample, due to its statistical effectiveness and clinical efficiency. Of the total sample of 512,

this cut-score yielded 59 false positives and 21 false negatives. Non-imputed results presented as imputed results were concordant.

Compared to the true negatives, the false positive cases on the BITSEA ASD-Competence scale, (cut-score of 11.5), had significantly lower scores on the following MSEL subscales: Visual Reception ($t(268)=3.45$, $p<.01$), Receptive Language ($t(266)=4.93$, $p<.001$), and Expressive Language ($t(266)=5.19$, $p<.001$). The false positives also obtained significantly lower MSEL ELC scores than the true negatives ($t(242)=3.80$, $p<.001$). Compared to the true positives, the false negative cases on the BITSEA ASD-Competence scale, (cut-score of 11.5), had significantly higher scores on the MSEL Receptive Language subscale ($t(199)=2.73$, $p<.01$), but they did not differ significantly on Expressive Language ($t(199)=1.28$, $p=.20$).

Discussion

In the current study, preliminary cut-scores were identified for the three newly derived subscales of the BITSEA: the ASD-Problem subscale, the ASD-Competence subscale, and the ASD-Total subscale. A sample of 223 toddlers, previously diagnosed with ASD using state of the art measurement, was compared to a sample of 289 toddlers with either significant non-ASD early-emerging psychopathology, developmental delays and disorders, or typical development. Unlike previous studies, the comparison sample in this study was enriched for psychological problems and diagnoses as well as developmental challenges, mirroring the kind of decision-making regularly encountered by specialty pediatric healthcare providers, such as developmental pediatricians and early interventionists. If used in primary care settings, it would be reasonable to expect specificity to be higher than observed in this study, as more children without ASD would be truly asymptomatic.

It was possible to identify cut-scores on each of the newly derived scales that achieved discrimination of children with and without ASD that support use of the BITSEA for this purpose. The following cut-scores were found to be the most effective for the ASD-Problem, ASD-Competence, and ASD-Total subscales respectively: greater than or equal to 4.5, less than or equal to 11.5, and greater than or equal to -7.5. These three cut-scores were found to have moderate to high discriminative power, as well as moderate to high sensitivity, specificity, and PPV. Given heightened interest in early detection (Zwaigenbaum et al., 2015), separate analyses of the subsample of children at or below 24 months revealed comparable, if not stronger, findings. Of note, the ASD-Competence subscale, which consists of questions regarding delays in the acquisition of typical social-communication and social-emotional competencies, outperformed the ASD-Problem and ASD-Total subscales in identifying both true positives and true negatives. This is consistent with the idea that delays or deficits in early social-communication and social-emotional behaviors distinguish children with ASD, rather than early symptoms of ASD such as repetitive behaviors. In addition, the finding that the ASD-Problem subscale is less effective than the ASD-Competence subscale demonstrates the challenges that caregivers have in distinguishing between age-appropriate repetitive and restricted behaviors and symptoms of ASD (e.g., repeating a behavior in the service of mastery versus perseverating in too consistent a

manner to indicate discovery, typical sensory exploration versus sensory-seeking behaviors, toddler persistence versus difficulty with transitions).

Consistent with our hypotheses, an examination of cognitive and linguistic characteristics associated with false positive status revealed that children who falsely screened positive had significantly lower cognitive and language abilities compared to the children with true negative status. This finding suggests that the specificity of this subscale is possibly limited by its overlap with symptoms of general intellectual or developmental disabilities or other early emerging psychopathology. Following screening, the next step in the process of evaluation, which may either consist of a secondary ASD-specific screener or a full diagnostic assessment, should include a test of cognitive and language abilities, as this would aid in differentiating between cognitive impairments and ASD symptomology. Additionally, hypotheses regarding the children who falsely screened negative were partially supported as the false negative cases had significantly higher receptive language abilities, although not expressive language abilities, compared to children who correctly screened positive. This finding suggests that advanced language abilities may conceal the specific ASD symptoms about which the screener inquires. While future research should explore using different cut-scores varying by a child's language abilities to determine risk for ASD, at this time using a single cut-score to optimize clinical utility of the scale is prioritized.

In order for screening tools to be effective, they not only need to have strong psychometric properties, but they also must be practical for use in healthcare settings and easily integrated into standing procedures. Because of the limited resources available to pediatric healthcare providers (e.g., time, support staff), screening tools should be short, easy for parents to understand, and quick and efficient to score (Barton et al., 2012). The BITSEA allows healthcare providers to detect ASD symptomology while simultaneously assessing other areas of a child's development and fulfilling the AAP and CDC's recommendation for regular developmental surveillance and screening of behavioral health. Moreover, use of the BITSEA ASD-Competence subscale, which was shown to have strong sensitivity and specificity within this sample, can be used in pediatric healthcare and early intervention settings to detect children who require further evaluation of ASD symptoms. Moreover, the use of positively worded questions, in addition to negatively worded questions, may be more acceptable to parents.

However, in interpreting the results of this study, some limitations should be kept in mind. First, the sample used in this study consisted of the samples from three previous projects and represent a convenience sample. While the participants in each study completed the BITSEA as part of a larger evaluation, none of the studies were specifically designed to develop ASD screening criteria on the BITSEA. The results presented capitalize on cut-scores that were developed in this convenience sample. Therefore, the external validity of the current study's results is uncertain until the procedure is replicated in an independent sample.

Moreover, because the prevalence rate of ASD was exponentially greater in this sample than in actuality, the sensitivity and specificity rates may change when the study is replicated in a general or high-risk sample. Increased base rates directly influence positive predictive value and can also bias estimates of sensitivity and specificity (Sheldrick et al., 2015). This further

emphasizes the need to replicate the current study in the general population to confirm the effectiveness of the cut-score scores on the various ASD subscales.

The findings of the current study contribute to the overarching goal of increased early detection and diagnosis of ASD and give pediatric healthcare providers and clinicians on the frontlines an efficient and effective ASD screening tool. This work sets the stage for testing these cut-scores in more diverse early intervention and pediatric settings. It will be important to examine the effectiveness of using the BITSEA for screening for ASD symptoms as well as other early-emerging social-emotional and behavioral concerns, compared to both ASD-specific screeners and other broadband behavioral screeners with separate ASD scales for toddlers. Future work should investigate whether or not parental reaction to completing a broad behavioral screener differs in terms of anxiety, compared to completing an ASD-specific screener. Previous research has documented the high level of stress that parents feel during the screening and diagnostic process (Siklos & Kerns, 2007), and therefore starting off globally with a broad-based screener may afford more time for the provider to introduce the idea of ASD to the parent. Similarly, anecdotal evidence demonstrates that in settings where universal screening for ASD is being implemented, parents are resistant to completing ASD screeners due to a lack of understanding of universal screening and a belief that they are being asked to complete the screener because of a suspicion that something is "wrong" with their child. Parents may be more willing to complete screening questionnaires if the measures were focused on broad behaviors, instead of narrowly focused on the symptoms of one disorder.

Overall, this study provides preliminary evidence for the effectiveness of the BITSEA as an effective screening tool for detecting toddlers at risk for ASD. The efficient nature of the BITSEA addresses the common concerns of pediatric healthcare providers and infant mental health providers related to the limit of time and resources within the field. Universal implementation of the BITSEA as a developmental screener in primary care and other healthcare settings has the power to reduce the disparities that are observed in age of diagnosis across various racial and ethnic groups, as well as the gap between first parental concern and diagnosis.

Acknowledgments

References

Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. Acta Paediatrica. 2007; 96(5):644–647. DOI: 10.1111/j1651-2227.2006.00178.x [PubMed: 17376185]

Akshoomoff N. Use of the mullen scales of early learning for the assessment of young children with autism spectrum disorders. Child Neuropsychology. 2006; 12(4-5):269–277. DOI: 10.1080/09297040500473714 [PubMed: 16911972]

Al-Qabandi M, Gorter J, Rosenbaum P. Early autism detection: Are we ready for routine screening? Pediatrics. 2011; 128(1):e211–e217. DOI: 10.1542/peds.2010-1881 [PubMed: 21669896]

Allison C, Baron-Cohen S, Wheelwright S, Charman T, Richler J, Pasco G, Brayne C. The Q-CHAT (Quantitative CHecklist for Autism in Toddlers): A normally distributed quantitative measure of autistic traits at 18-24-months of age: Preliminary report. Journal of Autism and Developmental Disorders. 2008; 38(8):1414–1425. DOI: 10.1007/s10803-007-0509-7 [PubMed: 18240013]

American Academy of Pediatrics, Council on Children with Disabilities, Section on Developmental and Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children with Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. Pediatrics. 2006; 118:405–420. DOI: 10.1542/peds.2006-1231 [PubMed: 16818591]

Baio, J. Morbidity and Mortality Weekly Report: Surveillance Summaries. Vol. 61. Centers for Disease Control and Prevention; 2012. Prevalence of Autism Spectrum Disorders: Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008.

Barbaro J, Dissanayake C. Autism spectrum disorders in infancy and toddlerhood: A review of the evidence on early signs, early identification tools, and early diagnosis. Journal of Developmental and Behavioral Pediatrics. 2009; 30(5):447–459. [PubMed: 19823139]

Baron-Cohen S, Allen J, Gillberg C. Can autism be detected at 18 months? The needle, the haystack, and the CHAT. The British Journal of Psychiatry. 1992; 161(6):839–843. [PubMed: 1483172]

Barton ML, Dumont-Mathieu T, Fein D. Screening young children for autism spectrum disorders in primary practice. Journal of Autism and Developmental Disorders. 2012; 42(6):1165–1174. [PubMed: 21842325]

Briggs-Gowan, MJ., Carter, AS. Examiner's Manual for the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). San Antonio, Texas: Psychological Corporation, Harcourt Press; 2006.

Briggs-Gowan MJ, Carter AS. Social-emotional screening status in early childhood predicts elementary school outcomes. Pediatrics. 2008; 121(5):957–962. DOI: 10.1542/peds2007-1948 [PubMed: 18450899]

Briggs-Gowan MJ, Carter AS, Irwin J, Wachtel K, Cicchetti DV. The Brief Infant-Toddler Social and Emotional Assessment: Screening for social-emotional problems and delays in competence. Journal of Pediatric Psychology. 2004; 29(2):143–155. [PubMed: 15096535]

Briggs-Gowan MJ, Carter AS, McCarthy K, Augustyn M, Caronna E, Clark R. Clinical validity of a brief measure of early childhood social-emotional/behavioral problems. Journal of Pediatric Psychology. 2013; 38(5):577–587. [PubMed: 23603252]

Bryson SE, Zwaigenbaum L, Brian J, Roberts W, Szatmari P, Rombough V, McDermott C. A prospective case series of high-risk infants who developed autism. Journal of Autism and Developmental Disorders. 2007; 37(1):12–24. DOI: 10.1007/s10803-006-0328-2 [PubMed: 17211728]

Dietz C, Swinkels S, van Daalen E, van Engeland H, Buitelaar JK. Screening for autistic spectrum disorder in children aged 14-15 months. II: Population screening with the Early Screening of Autistic Traits Questionnaire (ESAT). Design and General Findings. Journal of Autism and Developmental Disorders. 2006; 36(6):713–722. DOI: 10.1007/s10803-006-0114-1 [PubMed: 16633887]

Fluss R, Faraggi D, Reiser B. Estimation of the Youden's index and its associated cutoff point. Biometrical Journal. 2005; 47:458–472. DOI: 10.1002/bimj.200410135 [PubMed: 16161804]

Frohna JG. Failure to respond to name is indicator of possible autism spectrum disorder. The Journal of Pediatrics. 2007; 151(3):327–328. DOI: 10.1016/j.jpeds.2007.07.023 [PubMed: 17719953]

Gray KM, Tonge BJ. Screening for autism in infants and preschool children with developmental delay. Australian and New Zealand Journal of Psychiatry. 2005; 39(5):378–386. DOI: 10.1111/j.1440-1614.2005.01585.x [PubMed: 15860026]

Jones EH, Gliga T, Bedford R, Charman T, Johnson MH. Developmental pathways to autism: A review of prospective studies of infants at risk. Neuroscience and Biobehavioral Reviews. 2014; 39:1–33. DOI: 10.1016/j.neubiorev.2013.12.001 [PubMed: 24361967]

Kruizinga I, Jansen W, de Haan CL, van der Ende J, Carter AS, Raat H. Reliability and validity of the Dutch version of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). Plos ONE. 2012; 7(6)doi: 10.1371/journal.pone.0038762

Landa R, Garrett-Mayer E. Development in infants with autism spectrum disorders: A prospective study. Journal of Child Psychology and Psychiatry. 2006; 47:629–638. DOI: 10.1111/j. 1469-7610.2006.01531.x [PubMed: 16712640]

Lord C, Risi S, Lambrecht L, Cook ER, Leventhal BL, DiLavore PC, Rutter M, et al. The Autism Diagnostic Observation Schedule Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders. 2000; 30(3):205–223. DOI: 10.1023/A:1005592401947 [PubMed: 11055457]

Mandell DS, Listerud J, Levy SE, Pinto-Martin JA. Race differences in the age at diagnosis among Medicaid-eligible children with autism. Journal of The American Academy of Child & Adolescent Psychiatry. 2002; 41(12):1447–1453. DOI: 10.1097/00004583-200212000-0001 [PubMed: 12447031]

Matson, JL., Boisjoli, JA., Wilkins, J. Baby and Infant Screen for Children with aUtIsm Traits (BISCUIT). Baton Rouge, LA: Disability Consultants, LLC; 2007.

Mullen, EM. AGS Edition. Circle Pines, MN: American Guidance Service, Incorporated; 1995. Mullen Scales of Early Learning.

Ozonoff S, Iosif A, Baguio F, Cook IC, Hill M, Hutman T, Young GS, et al. A prospective study of the emergence of early behavioral signs of autism. Journal of the American Academy of Child & Adolescent Psychiatry. 2010; 49(3):256–266. DOI: 10.1097/00004583-201003000-00009 [PubMed: 20410715]

Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, Stone WL, et al. Recurrence risk for autism spectrum disorders: A Baby Siblings Research Consortium study. Pediatrics. 2011; 128(3):e488–e495. [PubMed: 21844053]

Pickles A, Le Couteur A, Leadbitter K, Salomone E, Cole-Fletcher R, Tobin H, Aldred C, et al. Parent-mediated social communication therapy for young children with autism (PACT): Long-term follow-up of a randomised controlled trial. The Lancet. 2016; 388:2501–2509.

Reznick J, Baranek GT, Reavis S, Watson LR, Crais ER. A parent-report instrument for identifying one-year-olds at risk for an eventual diagnosis of autism: The First Year Inventory. Journal of Autism and Developmental Disorders. 2007; 37(9):1691–1710. DOI: 10.1007/s10803-006-0303-y [PubMed: 17180716]

Robins DL, Fein D, Barton ML, Green JA. The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. Journal of Autism and Developmental Disorders. 2001; 31(2):131–144. DOI: 10.1023/A:1010738829569 [PubMed: 11450812]

Seida J, Ospina MB, Karkhaneh M, Hartling L, Smith V, Clark B. Systematic reviews of psychosocial interventions for autism: An umbrella review. Developmental Medicine and Child Neurology. 2009; 51(2):95–104. DOI: 10.1111/j1469-8749.2008.03211.x [PubMed: 19191842]

Sheldrick RC, Benneyan JC, Kiss IG, Briggs-Gowan MJ, Copeland W, Carter AS. Thresholds and accuracy in screening tools for early detection of psychopathology. Journal of Child Psychology and Psychiatry. 2015; 56(9):936–948. [PubMed: 26096036]

Sheldrick RC, Garfinkel D. Is a positive developmental-behavioral screening score sufficient to justify a referral? A review of evidence and theory Academic Pediatrics. in press.

Sheldrick RC, Merchant S, Perrin EC. Identification of developmental-behavioral problems in primary care: A systematic review. Pediatrics. 2011; 128(2):356–363. DOI: 10.1542/peds.2010-3261 [PubMed: 21727101]

Siegel, B. Pervasive Developmental Disorders Screening Test-II (PDDST-II). San Antonio, TX: Harcourt Assessment; 2004.

Siklos S, Kerns KA. Assessing the diagnostic experiences of a small sample of parents of children with autism spectrum disorders. Research in Developmental Disabilities. 2007; 28(1):9–22. DOI: 10.1016/j.ridd.2005.09.003 [PubMed: 16442261]

Swets J. Measuring the accuracy of diagnostic systems. Science. 1988; 240(4857):1285–1293. [PubMed: 3287615]

Swets JA, Dawes RM, Monahan J. Psychological science can improve diagnostic decisions. Psychological Science in the Public Interest. 2000; 1(1):1–26. [PubMed: 26151979]

Wetherby AM, Brosnan-Maddox S, Peace V, Newton L. Validation of the infant-toddler checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. Autism. 2008; 12(5):487–511. DOI: 10.1177/1362361308094501 [PubMed: 18805944]

Wetherby AM, Woods J, Allen L, Cleary J, Dickinson H, Lord C. Early indicators of autism spectrum disorders in the second year of life. Journal of Autism and Developmental Disorders. 2004; 34(5): 473–493. DOI: 10.1007/s10803-004-2544-y [PubMed: 15628603]

Woods JJ, Wetherby AM. Early identification of and intervention for infants and toddlers who are at risk for autism spectrum disorder. Language, Speech, and Hearing Services in Schools. 2003; 34(3):180–193. DOI: 10.1044/0161-1461(2003/015)

Yirmiya N, Gamliel I, Pilowsky T, Feldman R, Baron-Cohen S, Sigman M. The development of siblings of children with autism at 4 and 14 months: Social engagement, communication, and cognition. Journal of Child Psychology and Psychiatry. 2006; 47(5):511–523. DOI: 10.1111/j. 1469-7610.2005.01528.x [PubMed: 16671934]

Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3(1):32–35. [PubMed: 15405679]

Young RL, Brewer N, Pattison C. Parental identification of early behavioural abnormalities in children with autistic disorder. Autism. 2003; 7(2):125–143. DOI: 10.1177/1362361303007002002 [PubMed: 12846383]

Zhou H, Qin G. New nonparametric confidence intervals for the Youden's index. Journal of Biopharmaceutical Statistics. 2012; 22:1244–1257. DOI: 10.1080/10543406.2011.592234 [PubMed: 23075020]

Zwaigenbaum L, Bauman ML, Stone WL, Yirmiya N, Estes A, Hansen RL, Wetherby A, et al. Early identification of autism spectrum disorder: Recommendations for practice and research. Pediatrics. 2015; 136(4Suppl 1):S10–S40. [PubMed: 26430168]

Zwaigenbaum L, Bryson S, Lord C, Rogers S, Carter A, Carver L, Yirmiya N, et al. Clinical assessment and management of toddlers with suspected autism spectrum disorder: Insights from studies of high-risk infants. Pediatrics. 2009; 123(5):1383–1391. DOI: 10.1542/peds.2008-1606 [PubMed: 19403506]

Zwaigenbaum L, Bryson S, Rogers T, Roberts W, Brian J, Szatmari P. Behavioral manifestations of autism in the first year of life. International Journal of Developmental Neuroscience. 2005; 23:143–152. DOI: 10.1016/j.ijdevneu.2004.05.001 [PubMed: 15749241]
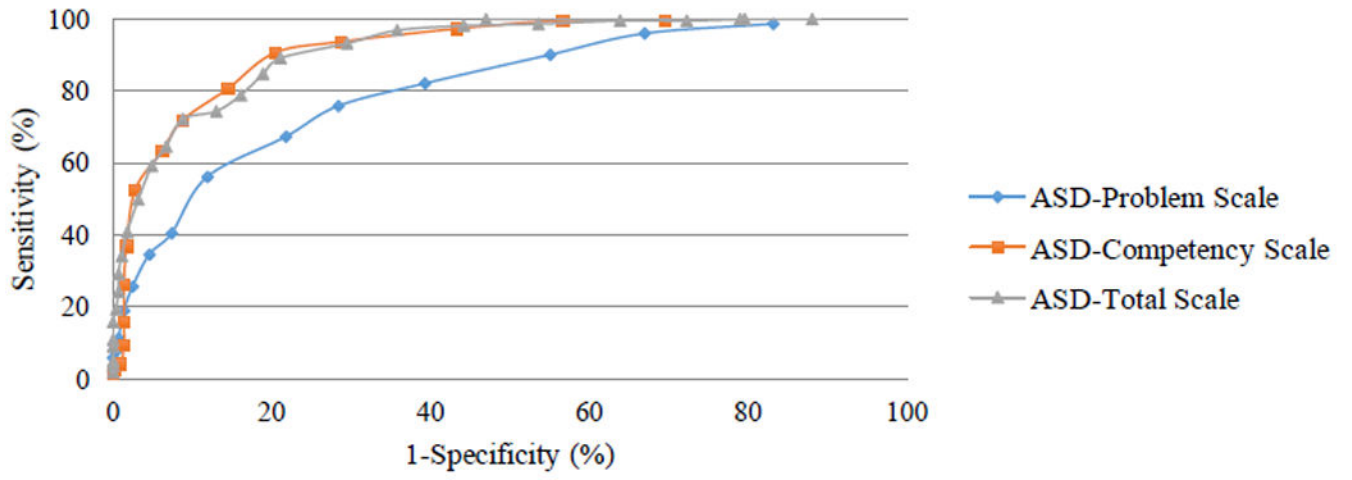
Figure 1.
ROC Plot of BITSEA ASD Subscales

Table 1

Behaviors Assessed on BITSEA ASD Subscales

Subscale	Behaviors Assessed[*]
ASD-Problems	Limited enjoyment of playful activities
	Unresponsive when hurt
	Tactile sensitivities
	Difficulty with transitions
	Repetitive play
	Repetitive speech
	Repetitive motor movements
	Appears unaware of surroundings
	Limited eye contact
	Avoidant of physical contact
ASD-Competence	Shares successes
	Seeks out caregiver when upset
	Responds to name with eye contact
	Emotional/physical affection
	Interactive play
	Empathy when someone is hurt
	Imitation skills
	Use of pointing to share interests/joint attention
	Pretend play

[*] Actual BITSEA items not provided due to copyright rules

Table 2

Participant Demographic Information

	ASD Sample		Comparison Sample
	Study 1	**Study 2**	**Study 3**
N	168	55	282
Mean Age in Months (SD)	27.9 (4.2)	21.3 (2.8)	31.0 (9.2)
% Male	78.5	54.3	64.1
SES	70% annual income > $60,000	Not reported	45% of families living below the federal poverty line
Parent Education Level	97% 2+ years of college	16% high school, 33% some college, 35% college degree, 16% advanced degree	27.3% some college, 42.6% college degree or greater
Race	83% White	47.4% White, 38.6% Latino, 5.5% Black, 5.3% Asian/White, 10.6% multiracial	45.5% White, 27.2% Black, 15% Latino, 6.6% multiracial, 5.7% other
Recruitment Sources	Early intervention providers, physicians specializing in ASD, local conferences, events for families of children with ASD	ASD specialty clinics, early intervention programs, pediatric and neurology practices, online Interactive Autism Network	Mental health and developmental assessments and services (pediatric clinics, home-visiting programs, early intervention programs)
Diagnostic Criteria	ASD was determined by DSM-IV criteria for autism or PDD-NOS, ADOS, and expert clinical judgment	ASD was determined by DSM-IV criteria for autism or PDD-NOS, ADOS, and expert clinical judgment	ASD was an exclusionary criterion for this study; videotapes of parent-child interactions were reviewed by licensed psychologists and developmental pediatricians who were familiar with the presentation of ASD in early childhood

Table 3

Comparison of Cut-scores Across Scales

Scale	Cut-score	FN	TN	FP	TP	Sensitivity (95% CI)	Specificity (95% CI)	AUC	Youden's Index (*J*)
Full sample									
ASD-Problems	4.5	54	204	81	169	75.8% (69.5-81.1%)	71.6% (65.9-76.7%)	0.81	0.47
ASD-Competence	11.5	21	230	59	202	90.6% (85.8-93.9%)	79.6% (74.4-84.0%)	0.92	0.70
ASD-Total	-7.5	24	225	60	199	89.2% (84.2-92.8%)	78.9% (73.7-83.4%)	0.92	0.68
Subsample of children 24 months and younger									
ASD-Problems	4.5	25	58	19	64	71.9% (61.2-80.7%)	75.3% (64.0-84.1%)	0.80	0.47
ASD-Competence	11.5	4	59	20	85	95.5% (88.3-98.6%)	74.7% (63.4%-83.5%)	0.94	0.70
ASD-Total	-7.5	8	62	15	81	91.0% (82.6-95.8%)	80.5% (69.6-88.3%)	0.92	0.72