



Published in final edited form as:

Autism. 2018 October ; 22(7): 881–890. doi:10.1177/1362361317712649.

Utilizing two-tiered screening for early detection of autism spectrum disorder

Meena Khowaja,

Department of Psychology, Georgia State University, USA

Diana L Robins, and

AJ Drexel Autism Institute, Drexel University, USA

Lauren B Adamson

Department of Psychology, Georgia State University, USA

Abstract

Despite advances in autism screening practices, challenges persist, including barriers to implementing universal screening in primary care and difficulty accessing services. The high false positive rate of Level 1 screening methods presents especially daunting difficulties because it increases the need for comprehensive autism evaluations. The current study explored whether two-tiered screening – combining Level 1 (Modified Checklist for Autism in Toddlers, Revised with Follow-Up) and Level 2 (Screening Tool for Autism in Toddlers & Young Children) measures – improves the early detection of autism. This study examined a sample of 109 toddlers who screened positive on Level 1 screening and completed a Level 2 screening measure prior to a diagnostic evaluation. Results indicated that two-tiered screening reduced the false positive rate using published STAT cutoffs compared to Level 1 screening alone, although at a cost to sensitivity. However, alternative STAT scoring in the two-tiered screening improved both positive predictive value and sensitivity. Exploratory analyses were conducted, including comparison of autism symptoms and clinical profiles across screening subsamples. Recommendations regarding clinical implications of two-tiered screening and future areas of research are presented.

Keywords

Autism; screening; early detection; M-CHAT; M-CHAT-R; STAT

Introduction

Implementing screening for autism spectrum disorder (ASD) and other developmental delays in primary care settings can improve early identification and lower the age at which children are referred for necessary intervention (Pinto-Martin et al., 2005; Guevara et al., 2013), which can enhance prognostic outcome (MacDonald et al., 2014; Orinstein et al., 2014; Reichow, 2012; Reichow et al., 2012). However, many barriers persist in

implementing standardized screening and evaluation referral, including time constraints in primary care settings and difficulty identifying accurate and feasible tools. The current study evaluates the effectiveness of two-tiered screening to identify toddlers at risk for ASD, using the Modified Checklist for Autism in Toddlers, Revised with Follow-Up (M-CHAT-R/F; Robins et al., 2009) as a Level 1 screener and the Screening Tool for Autism in Toddlers & Young Children (STAT; Stone and Ousley, 1997) as a Level 2 screener.

The American Academy of Pediatrics (AAP) guidelines for early detection of ASD (Johnson et al., 2007) encompass three complementary approaches: (1) ongoing developmental surveillance, (2) broad developmental screening at 9, 18, and 24/30 months, and (3) universal ASD-specific screening at 18- and 24-month visits. Children at risk for ASD should be referred for diagnostic evaluation. Level 1 screeners are designed to differentiate toddlers at-risk for ASD from typically developing children in the general primary care population, whereas Level 2 tools are more often used in early childhood developmental programs to distinguish ASD risk from risk for other developmental disorders (Johnson et al., 2007). Level 1 measures usually are parent questionnaires. These can be convenient and cost-effective to administer, though they are dependent upon the parents' knowledge and awareness of particular constructs related to ASD and normative child development (for review, see Barton et al., 2012a). Some Level 2 ASD screeners are parent questionnaires, but many rely on a trained clinician's direct observation of the presence or absence of ASD-related behaviors that parents might not readily recognize (for review, see Norris and Lecavalier, 2010). Children identified as at-risk for ASD via ongoing developmental surveillance (e.g. sibling with ASD, parent concern) should be screened with the appropriate Level 1 or 2 tools (Johnson et al., 2007).

Estimates of pediatrician screening practices are variable, ranging from 22% to 82% for broad developmental screening (Dosreis et al., 2006; Gillis, 2009; Pierce et al., 2011; Radecki et al., 2011) and 8% to 60% for ASD-specific screening (Arunyanart et al., 2012; Dosreis et al., 2006; Gillis, 2009; Self et al., 2015); importantly, pediatrician screening rates are increasing (Radecki et al., 2011). Barriers have included limited time and resources to screen (Carbone et al., 2010; King et al., 2010), lack of knowledge about ASD-specific screening tools (Dosreis et al., 2006), and reduced confidence identifying symptoms (Zuckerman et al., 2013). Even when screening is conducted, pediatricians may not refer at-risk cases for follow-up (King et al., 2010), suggesting they may doubt that screening tools effectively identify ASD risk (Barton et al., 2012a). These barriers persist and indicate a continued need for effective, feasible screening tools and procedures to help identify children who warrant referrals to specialists.

Because no screening tool has perfect sensitivity and specificity, it is necessary to consider the implications of false negative (FN; child with ASD screens *not* at risk) and false positive (FP; non-ASD child screens at risk) results. In the context of this study, FP and FN refer specifically to ASD detection, rather than the need for referral for diagnostic testing for any developmental delay. Children with developmental or language delays other than ASD often screen positive on Level 1 ASD screeners. In an earlier study (Robins et al., 2014), it was found that 95% of toddlers who screen positive on the M-CHAT-R/F were diagnosed with some type of delay, approximately half of which were ASD (Robins et al., 2014). Thus, FP

indicates a child who was referred for evaluation after screening positive and did not receive an ASD diagnosis, which implies that the child could benefit from referral to professionals who are not experts in ASD and could address non-ASD related developmental concerns.

Disorders like ASD, in which the timeliness of intervention onset impacts long-term prognosis, require minimal FN cases (Harris and Handleman, 2000; Perry et al., 2011). If the majority of children who screen positive were truly in need of differential diagnostic evaluations for ASD (i.e. minimizing FNs and FPs), this could reduce the burden on evaluation and intervention referral process. However, utilizing only one Level 1 screening tool may not be optimal in achieving this differentiation. One important strategy to consider is utilizing a two-tiered screening procedure in which children who screen positive on the Level 1 screener are screened again using a Level 2 screening before being referred for a diagnostic clinical evaluation.

Implementation of two-tiered screening has been effective for many disorders, including gestational diabetes (Meltzer et al., 2010), hepatocellular carcinoma (Shih et al., 2010), trisomy 21 (Nicolaidis et al., 2005), and sleep-disordered breathing (Chen et al., 2011). But to date two-tiered screening for ASD has rarely been used. In the Netherlands, investigators explored using the Early Screening of Autistic Traits (ESAT), a yes/no questionnaire designed for screening in the general population that can be completed by parents or by professionals with parent interview (Dietz et al., 2006; Swinkels et al., 2006), as two-tiered protocol for ASD screening on a large scale ($n=31,724$) in an urban province of 1.1 million residents. However, they used the same instrument at both levels, with the initial pre-screen consisting of the first 4 items of the ESAT completed by parents during primary care visits, followed by the full 14 items as the Level 2 screener administered during subsequent home visits. Of the 73 14- to 15-month-olds who screened positive for ASD-risk after the second screen and were evaluated, only 18 were diagnosed with ASD, yielding a PPV of .25; all FPs had other developmental concerns. Psychometric properties (i.e. sensitivity, specificity) evaluating the ESAT's utility as a combined Level 1 and Level 2 screening protocol were not obtained, as children who screened negative on the initial prescreen were not followed up. Moreover, this two-stage protocol utilized sections from the same measure at both time points, rather than employing a multi-method approach. Two-tiered ASD screening is not standard practice in the United States and there is insufficient empirical support for the claim that a two-tiered screening approach maximizes detection of ASD while reducing the FP rate without increasing pediatrician burden. However, a recent initiative in South Carolina has demonstrated the promise of such multilevel ASD screening as a means to guide referrals for state-funded early intervention services (Rotholz et al., 2017). Specifically, screening positive on M-CHAT (Level 1) followed by subsequent screen positive on STAT (Level 2) results in "presumptive eligibility" for early intervention prior to having completed a diagnostic evaluation. This method has resulted in 5 times as many children referred for early intervention via presumptive eligibility, only 2.5% of whom were later found to not have an ASD diagnosis after evaluation (Rotholz et al., 2017). Although this referral process yielded higher rates of service utilization, an assessment of the psychometric statistics of two-tiered screening is needed to evaluate more fully the potential costs and benefits of implementing this relatively novel approach.

Current study

The goal of this study is to evaluate a two-tiered screening protocol combining two ASD-specific screening tools, the M-CHAT-R/F and the STAT. The M-CHAT-R/F is among the most widely used Level 1 screening tools, and has high sensitivity and specificity in low-risk settings (Robins et al., 2014). However, its moderate positive predictive value (PPV; .48) indicates that the FP rate for ASD is quite high, even though most of the children who screen positive have a significant developmental delay or concerns (PPV=.95-.98; Chlebowski et al., 2013; Robins et al., 2014). A high false positive rate can result in over-referrals for ASD diagnostic evaluations, which may lengthen waiting lists and put a strain on the system; this study seeks to explore whether secondary screening can help tailor referrals for ASD concerns vs. other developmental concerns.

To reduce the FP rate for ASD, children who screened positive on the M-CHAT-R/F were observed using the STAT, a structured, interactive play-based Level 2 screening measure administered by a trained clinician (see Figure 1). Both the M-CHAT-R/F and STAT address core domains of behaviors related to ASD symptoms, such as use of gestures or other nonverbal communication, imitative behavior, play skills, and joint attention behavior, but they do so using complementary methods of parent report of characteristic behavior and clinician observation using a structured protocol. This approach differs from two-stage screening with the ESAT in that it utilizes a multi-informant method with distinct measures at each level. We hypothesized that STAT screening as an intermediate step between positive M-CHAT-R/F and ASD diagnostic evaluation will reduce the FP rate (improve PPV) without compromising the combined sensitivity of these measures. If two-tiered screening for ASD is efficacious, it could be a powerful screening system that can better determine type of referrals that will best fit a child's needs (e.g. ASD-specific vs. general development or speech/language services) and inform state-level policy on the early detection and intervention of ASD.

Several exploratory analyses were also conducted to investigate how best to adapt the STAT for use in two-tiered screening, including evaluating alternative scoring criteria. Clinical outcomes were compared when using published versus alternative scoring methods as well as across those who did and did not screen positive on the STAT using published scoring criteria.

Method

Participants

The current study is part of a larger, ongoing community-based early screening project. All participants were enrolled in the large ($n=14,170$) screening study, in which 49 pediatric offices in metropolitan Atlanta conducted screening with the M-CHAT-R/F during well-child visits. Eligibility criteria consisted of 16- to 30-month-old children at initial M-CHAT-R/F screening whose parents were English-speaking and consented to the study. Exclusion criteria included severe sensory or motor impairment that interfered with evaluation. A subset of families in the screening study were asked if they were interested in participating

in optional, affiliated research studies; participating in studies that included completion of the STAT entailed a separate visit to prior to the evaluation.

The study sample ($n=109$) included all children who (a) screened positive on the 2-stage M-CHAT-R/F, (b) completed a STAT during an optional research study, and (c) completed a diagnostic evaluation. All participants in the study sample screened positive on the M-CHAT-R/F with the exception of one participant who bypassed Follow-Up after a high initial score on the M-CHAT-R.

Measures

The Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT-R/F; Robins et al., 2009; available at www.mchatscreen.com) is a Level 1 ASD screening tool. Parents complete 20 Yes/No items in approximately 5 minutes; if three or more items indicated risk, parents complete the structured Follow-Up. A total Follow-Up score of two or more indicated risk for ASD. Internal consistency (Cronbach's alpha) for the M-CHAT-R/F is 0.79, sensitivity is .85, specificity is .99, PPV is .48, and NPV is .99 (Robins et al., 2014).

The Screening Tool for Autism in Toddlers & Young Children (STAT; Stone and Ousley, 1997) is a 12-item Level 2 interactive screening tool designed for use with at-risk toddlers. Administration takes approximately 20 minutes. STAT administration took place at a university research lab with trained research staff who achieved 80% reliability on three observations and three administrations.

The STAT assesses four behavior domains: play, requesting, directing attention, and imitation. Each domain is scored as a proportion of items failed with 1 being the highest score (e.g. 3 failed imitation items = .75). Total score is the sum of domain scores (range: 0 to 4). Sensitivity, specificity, PPV, and NPV were reported to be .92, .85, .86, and .92, respectively, when administered to a validation sample of children already diagnosed with Autistic Disorder or other developmental delays or language impairments (Stone et al., 2004). Although the STAT was originally designed for children 24–35 months, it was shown to distinguish children with ASD from non-ASD in 14- to 23-month-old children (Stone et al., 2008), using a threshold of 2.75 (sensitivity .93, specificity .83, PPV .68, NPV .97) as opposed to 2.00 for older children.

A diagnostic evaluation was conducted with toddlers demonstrating risk for autism. Measures included the Mullen Scales of Early Learning (Mullen, 1995), Vineland Adaptive Behavior Scales-II (Sparrow et al., 2005), Behavioral Assessment System for Children-2 (Reynolds and Kamphaus, 2004), Autism Diagnostic Interview, Revised (Lord et al., 1994) or Toddler ASD Symptom Interview (Barton et al., 2012b), Childhood Autism Rating Scale-2 (Schopler et al., 2010), Autism Diagnostic Observation Schedule, first and second editions (ADOS(-2); Lord et al., 1999, 2012a, 2012b), and parent report of developmental history. ADOS severity scores were computed for the total score and for the Social Affect (SA) and Restricted and Repetitive Behaviors (RRB) domains (Esler et al., 2015; Hus et al., 2014).

Procedures

Parents completed either a paper or electronic version of the M-CHAT-R at the pediatrician's office. University research staff called parents whose child screened positive on the paper version to complete the Follow-Up by phone; those who completed the electronic M-CHAT-R were prompted to complete an electronic Follow-Up in the same session (see Table 1). Parents of children who continued to screen positive were invited to the university for a free diagnostic evaluation. Families were also asked if they would participate in additional research, which included participating in Level 2 screening (STAT) during a separate visit to the university prior to the evaluation appointment. STATs were administered by supervised trainees or research staff at the university, all of whom established reliable administration of the STAT.

During the diagnostic evaluation, a licensed psychologist and a graduate student clinician or research staff completed autism diagnostic measures and measures of cognitive, motor, language, and adaptive functioning. To minimize examiner bias, the evaluation team was blind to screening results. Using all available evaluation data and DSM-IV criteria, clinicians classified children in one of the following non-overlapping groups: ASD, other language or developmental disorders, no diagnosis (one or more scores outside typical range, but no applicable diagnosis), or typical development.

Data analysis

Data were analyzed using IBM SPSS Statistics 22.0 software. Psychometrics (i.e. sensitivity, specificity, PPV, NPV) for combined M-CHAT-R/F and STAT screening were computed. Receiver Operating Characteristic (ROC) analysis was conducted to determine accuracy of two-tiered screening in distinguishing ASD from non-ASD, as indicated by the area under the curve (AUC).

Alternative STAT scores were determined via Discriminant Function Analysis (DFA), which equally weighted all 12 STAT items. Items with the greatest discriminatory power (i.e. highest standardized canonical coefficients) constituted the alternative total score. The study sample was divided into two samples of equal sizes ($n=54$, $n=55$) matched on the child's diagnosis (58–59% ASD), sex (38% female), and age at time of STAT ($M=24$ months). ROC analysis was conducted on the first half of the sample to determine optimal scoring cutoffs for the alternative total score, and the alternative scoring was then applied to the second half of the sample. ASD symptom severity, cognitive ability, and adaptive functioning were related to screening outcomes (both original and alternative), using bivariate Pearson correlations. Finally, multivariate analyses of variance and Games-Howell post-hoc comparisons were conducted to compare clinical profiles across subsamples based on STAT results and diagnostic outcome.

Results

Using the original published scoring cutoffs for the STAT (i.e. 2.00 for children ≥ 24 months and 2.75 for children < 24 months; Stone et al., 2008), 30 of the 39 STAT screen positive children received an ASD diagnosis (See Figure 2), yielding a PPV of .77. Sensitivity,

specificity, and NPV of the STAT were .47, .80, and .51, respectively. The ROC analysis demonstrated adequate accuracy (AUC=.73), based on threshold of .70. This was a significant improvement in PPV compared to .475 for M-CHAT-R/F alone in Robins et al. (2014; $\chi^2(1, N=260)=11.49, p<.001$), but sensitivity declined significantly from .85 reported by Robins et al. (2014; $\chi^2(1, N=187)=31.07, p<.001$).

DFA was conducted to explore whether alternative STAT scoring methods demonstrate higher sensitivity. Standardized canonical coefficients indicated seven items – two directing attention, two play, two requesting, and one imitation (see Table 2) – had the strongest discriminatory power for ASD diagnosis. An alternative STAT score based on these seven items was calculated. ROC analysis using half the sample determined an optimal threshold of three maintained sensitivity .8 (see Figure 3), yielding good accuracy (AUC=.81). Applying this cutoff resulted in sensitivity of .81, specificity of .59, PPV of .74, and NPV of .68. This alternative scoring was applied to the validation sample; results were highly consistent: sensitivity=.78, specificity=.57, PPV=.71, and NPV=.65. Alternative scoring reduced the FN cases by 62% compared to published cutoffs, which significantly improved sensitivity, $\chi^2(1, N=64)=21.0, p<.001$. Also noteworthy is that all 30 children with ASD who screened positive on the STAT based on published cutoffs screened positive using the alternative total score.

The relationship between STAT scores and ADOS performance, cognitive testing, and parent report of adaptive functioning was examined to explore the relation between the child's performance during screening and the clinical evaluation. Both the original and alternative STAT scores were highly correlated with ADOS Total and SA severity scores and moderately correlated with the RRB severity score (see Table 3). STAT scores also had strong negative correlations with Mullen Early Learning Composite and Vineland-II Adaptive Behavior Composite.

Clinical Profiles Across STAT Screening Subsamples

Clinical phenotype of ASD cases was compared across children who had screened positive on the STAT and those who were STAT negative; see Table 4. Independent samples *t*-tests with Sidak-Bonferroni correction ($\alpha=.010$) indicated that among ASD cases, children who were STAT positive demonstrated significantly higher ADOS total ($t(62)=2.96, p=.004$) and SA ($t(62)=3.29, p=.002$) severity scores compared to STAT negative ASD cases; a similar direction of results was found for the RRB severity score ($t(62)=2.62, p=.011$), but was not significant. Additionally, both cognitive ($t(46.1)=-3.75, p<.001$) and adaptive ($t(62)=-3.49, p=.001$) functioning were found to be significantly lower for the STAT positive cases compared to the STAT negative cases ($ps<.01$). In comparing ADOS severity scores within the non-ASD cases across the two subsamples, no significant differences were found for ADOS severity scores, cognitive functioning, or adaptive functioning ($ps>.01$).

Discussion

This study's primary aim was to investigate whether implementing a two-tiered screening approach, in which the STAT is administered after children demonstrate risk on the M-CHAT-R/F, improves accuracy in detecting ASD. Using published scoring cutoffs, the STAT

reduced the FP rate compared to the rate when evaluation occurred immediately after a positive screen on the M-CHAT-R/F only, but at a cost to sensitivity. However, this tradeoff was reduced when an alternative STAT scoring algorithm was used. Moreover, all children with ASD who had screened positive on the STAT using the published scoring criteria also screened positive using the alternative scoring, indicating that detection of true positive cases was not compromised.

These results provide empirical support for two-tiered screening to streamline referrals for ASD-specific comprehensive evaluation or early intervention. With the intention of improving screening accuracy by minimizing the number of FN and FP cases, this study prioritized psychometric values of PPV and sensitivity. An important consideration in selection of screening methodology is the context of the referral and the tradeoff of psychometric values. For example, when the goal of a referral is to identify as many children with ASD as possible in order to initiate necessary early intensive behavioral interventions, then sensitivity may be of greater priority to help reduce missed cases. Such approach typically yields higher FPs, but the majority of these children are often identified as having non-ASD developmental delays that also require early intervention. However, resources for ASD evaluations are limited, and with more FP cases, the longer the waiting lists to see specialists.

Our finding that higher scores on the STAT, both using published and alternative scoring methods, were related to higher autism symptom severity and lower cognitive functioning and adaptive functioning among children with ASD is consistent with the literature on clinical profiles of children with ASD (Charman et al., 2011; Paul et al., 2014; Ray-Subramanian et al., 2011; Stone et al., 2004). Whereas significant differences in ADOS total and SA severity scores were found across STAT performance in children with ASD, this was not the case for the RRB severity score. This is to be expected, as the STAT items emphasize abnormalities in social communication and do not directly measure RRBs. Our results suggest that children who are more severely affected by ASD and have more impaired cognitive and adaptive functioning tend to screen positive across informants and methods.

The low sensitivity of the original published STAT cutoffs (Stone et al., 2004; Stone et al., 2008) in this study was surprising. Although developing new scoring criteria for the STAT was not the main purpose of this study, alternative scoring was examined, given that our sample is qualitatively quite distinct from samples used to develop these cutoffs. In their first study, Stone and colleagues (2004) used a sample of two-year-old children who were already diagnosed with developmental delays or Autistic Disorder, excluding children with PDD-NOS. In their subsequent study focusing on using the STAT in children younger than two years (Stone et al., 2008), PDD-NOS was included, but the sample was primarily composed of children who had an older sibling diagnosed with ASD. There may be qualitative or phenotypic differences in autism symptom presentation for children identified as at-risk for autism based on a known family history compared to children identified as at-risk based on a screening questionnaire administered to the general population. For example, Taylor and colleagues (2015) found that young children with ASD from multiplex families demonstrate greater social and pragmatic language impairment than those from simplex families. In addition to differences in symptom presentation, parents who already have experience

rearing a child with ASD may respond differently than those who do not on screening questionnaires.

Clinical Implications

The M-CHAT-R/F is the most widely-used screening measure to detect risk for ASD, but results in a high number of FPs. Although many of the FPs have other developmental delays and could benefit from early intervention referrals that are not ASD-specific, over-referrals for ASD diagnostic evaluations, which are costly and time-consuming for families, can lengthen waiting lists for evaluations and intervention. In the present study, we sought to investigate whether secondary ASD screening could streamline referrals to determine eligibility for ASD-specific services. Our results indicate that two-tiered screening, using the alternative scoring of the STAT, may more accurately detect ASD risk. PPV for ASD increased by 48% (M-CHAT-R/F alone PPV=.48; M-CHAT-R/F + STAT alternate scoring PPV=.71), with sensitivity decreasing by only 8% (M-CHAT-R/F only sensitivity=.85; M-CHAT-R/F + STAT alternate sensitivity=.78). Use of alternative scoring was used to explore how to refine the screening, evaluation, and intervention referral process for toddlers at risk for ASD. However, there are several additional areas related to multilevel ASD screening that require further research and consideration in order to inform any specific clinical recommendations.

A first practical challenge to be addressed in further studies is determining how secondary screening can be implemented in the community setting using accurate and efficient methods. For example, it may be difficult to accomplish secondary screening in the fast-paced setting of a child's primary care physician's office. Thus, it may be more feasible for children to be referred to clinics that specialize in autism evaluations to conduct Level 2 screening. The STAT, for example, requires less intensive training than many diagnostic tools, such as the ADOS-2, and is appropriate for service providers of varying backgrounds (e.g. undergraduate volunteers, graduate trainees, intake counselor, clinical social worker, nurse, research staff) to administer. The short administration time would allow several screening appointments to be completed within one day. This could reduce the number of children requiring comprehensive evaluation and allow families to be seen for evaluation more quickly. If replication studies utilizing the STAT (or other Level 2 measures) for secondary ASD screening show strong psychometric outcomes, this may support use of two-tiered screening for streamlining referral process.

Secondly, research on multilevel ASD screening design requires attention to the demands, costs, and benefits of different models for triaging and referral. Given the obstacle of long waiting lists for costly evaluations, two-tiered screening may potentially be used to prioritize children who more urgently need (i.e. screen positive on both measures) autism-specific assessment versus those who would benefit continued monitoring or more general developmental evaluation. Alternatively, as has been implemented in South Carolina (Rotholz et al. 2017), multilevel ASD screening may also be used to directly refer for government-funded early intensive behavioral interventions while waiting for an evaluation, given that children who screen positive on both steps of the two-step screening have a high likelihood of receiving an ASD diagnosis and demonstrating more severe symptoms.

Although this approach prioritizes autism evaluation referrals without delaying early intervention services, thereby streamlining the referral process via state policy changes, there may be barriers in other states that might hamper such a policy initiative.

Finally, there are important systemic factors impacting the implementation of multilevel ASD screening that require consideration and improvement, including insurance coverage and public policies. Ideally, secondary screening would be covered under recent insurance mandates for ASD, as would coverage of ASD-specific intervention services delivered based on screening positive on tools like the STAT rather than delaying service delivery until a formal diagnosis is made. However, the extent to which screening, diagnosis, and treatment are covered by insurance varies widely by state. Therefore, it is important to consider types of public early interventions offered and breadth of insurance coverage in regard to implementation of two-tiered screening. For example, public early intervention services may typically include speech, physical, and occupational therapy, which are certainly beneficial for many young children with ASD, but ASD-specific behavioral interventions (e.g. Applied Behavioral Analysis) are not always embedded in public programs, or available to children who do not have a formal ASD diagnosis. With recent mandates, such behavioral interventions can sometimes be offered via insurance, but only after a diagnosis has been made. Continued efforts are needed to not only help expand insurance coverage and eligibility for intervention services, but also increase government funding for providing evidence-based, ASD-specific early intensive behavioral interventions.

Limitations

Although this study conducted initial screening in community-based pediatric clinics, secondary screening and evaluation took place in a university setting following a research protocol. As such, future research should investigate the clinical utility of the proposed two-tiered screening methodology within community practices. In our sample, the model was strongest when using modified STAT cutoffs; however, it was not the aim of this study to develop a new STAT scoring protocol, but rather to focus on exploring the efficacy of a two-stage screening model. Additionally, the sample was divided in half for development and validation of the alternative STAT cutoffs, resulting in small sample sizes. Therefore, cross-validation of the modified STAT scoring is needed in larger and independent samples. Additional studies that expand on utilizing a two-step approach with Level 1 and Level 2 measures also will be beneficial.

Moreover, it is important to consider ways to streamline the flow between the various steps in the screening and evaluation process. The more steps needed before clinical evaluation, the more children may be lost. In the current study, attrition limits the interpretation and generalization of the findings regarding two-tiered screening. More generally, attrition is a daunting concern in community-based screening research (e.g. Janvier et al., 2016; Pierce et al., 2011). Based on estimates from the larger screening project from which our sample was drawn, 21% of families with screen positive M-CHAT or M-CHAT-R results did not complete the Follow-Up. Of the 442 participants who screened positive on the M-CHAT-R/F, warranting further evaluation, 24% completed both the STAT and the evaluation (sample in current study), 32% completed the evaluation only, 2% completed the STAT only,

and 42% did not complete the STAT nor the evaluation. Therefore, attrition between Level 1 screening and diagnostic evaluation was 44%. Attrition in the larger screening study was correlated with lower maternal education and racial/ethnic minority status; the most common reasons for not participating in the evaluation were non-responsiveness when contacted (i.e. not attending the scheduled appointment, or not responding to scheduling calls) and parents declining to participate when offered (see Khowaja et al., 2015). It is likely that there may be attrition as well between Level 1 and 2 screening, although our study did not address this issue because the STAT was performed as part of an optional research visit that took place prior to the diagnostic evaluation and not as part of the larger screening study protocol. However, it is noteworthy that of the 117 parents who participated in the STAT visit, most (93%) subsequently completed the evaluation, suggesting low attrition from Level 2 screening to evaluation. Also, only 2% of those who screened positive on Level 1 screening completed the STAT but did not complete the evaluation, indicating that only rarely did participation in Level 2 screening hinder follow-through with diagnostic evaluation. Nevertheless, although Level 2 screening may help streamline referrals and reduce the FP rate, it may also be less convenient for families. For example, some families may find it challenging to schedule a visit at a specialty clinic for Level 2 screening. Thus, in implementing two-step screening, increased outreach efforts from the clinic may be necessary to help parents of varying sociodemographic backgrounds navigate towards thorough, accurate diagnosis.

Overall, findings from the current study are promising in terms of improving screening accuracy without increasing pediatrician burden. The use of two-tiered screening demonstrates a lower FP rate and, with alternative scoring, also demonstrates a lower FN rate. Additionally, it may be a cost-effective approach that streamlines referrals for evaluations and/or intervention needs. The combined M-CHAT-R/F and STAT methodology in this study is one approach to mitigating long waiting lists for diagnostic evaluations, one of the most pressing challenges that affect families and providers when a young child is identified as being at-risk for ASD. Future research should focus on identifying similar solutions, whether replicating use of M-CHAT-R/F and STAT, utilizing other measures in a two-tiered approach, or examining other novel approaches. Simultaneously, continued efforts are needed from scientists, legislators, providers, and families to advocate for policy changes to increase funding to provide accessible resources.

Acknowledgments

The authors gratefully acknowledge funding from the Eunice Kennedy Shriver National Institute of Child Health & Human Development [R01HD039961] and Autism Speaks Targeted Research Award [8368]. We also thank the participating families and the Developmental Neuropsychology Lab at Georgia State University for assistance with data collection, especially Ciara Braun, Bianca Brooks, Courtney Cadle, Karis Casagrande, Liz Dlouhy, Deborah Fein, Sarah Fink, Ann Grossniklaus, Kiauhna Haynes, Anquanette Herndon, Ashleigh Hover, Ashleigh Kellerman, Lauren Kendall, Molly Locklear, Makeda Moore, Chandler Puhly, Allison Ramsay, Riane Ramsey, Lauren Schmuck, Brandi Smith, Lauren Stites, Charu Subramanian, Katharine Suma, Janice Taylor, Oshianna Thomas, Ashley Tolleson, Lisa Wiggins, Amber Wimsatt, and Shelly Zody.

References

Arunyanart W, Fenick A, Ukritchon S, et al. Developmental and autism screening: A survey across six states. *Infants & Young Children*. 2012; 25(3):175–187.

- Barton M, Dumont-Mathieu T, Fein D. Screening Young Children for Autism Spectrum Disorders in Primary Practice. *Journal of Autism and Developmental Disorders*. 2012a; 42(6):1165–1174. [PubMed: 21842325]
- Barton M, Boorstein H, Herlihy L, et al. Toddler ASD Symptom Interview. 2012b Self-published.
- Carbone P, Behl D, Azor V, et al. The Medical Home for Children with Autism Spectrum Disorders: Parent and Pediatrician Perspectives. *Journal of Autism and Developmental Disorders*. 2009; 40(3): 317–324. [PubMed: 19768528]
- Charman T, Pickles A, Simonoff E, et al. IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). *Psychological Medicine*. 2011; 41(03):619–627. [PubMed: 21272389]
- Chen N, Chen M, Li H, et al. A two-tier screening model using quality-of-life measures and pulse oximetry to screen adults with sleep-disordered breathing. *Sleep and Breathing*. 2011; 15(3):447–454. [PubMed: 20449670]
- Chlebowski C, Robins DL, Barton ML, et al. Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics*. 2013; 131(4):e1121–e1127. [PubMed: 23530174]
- Dietz C, Swinkels S, van Daalen E, et al. 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. [PubMed: 16633887]
- Dosreis S, Weiner C, Johnson, et al. Autism Spectrum Disorder Screening and Management Practices Among General Pediatric Providers. *Journal of Developmental & Behavioral Pediatrics*. 2006; 27(Supplement 2):S88–S94. [PubMed: 16685190]
- Esler AN, Bal VH, Guthrie W, et al. The autism diagnostic observation schedule, toddler module: Standardized Severity Scores. *Journal of Autism and Developmental Disorders*. 2015; 45(9):2704–2720. [PubMed: 25832801]
- Gillis J. Screening Practices of Family Physicians and Pediatricians in 2 Southern States. *Infants & Young Children*. 2009; 22(4):321–331.
- Guevara J, Gerdes M, Localio R, et al. Effectiveness of Developmental Screening in an Urban Setting. *Pediatrics*. 2012; 131(1):30–37. [PubMed: 23248223]
- Harris SL, Handleman JS. Age and IQ at intake as predictors of placement for young children with autism: A four-to six-year follow-up. *Journal of autism and developmental disorders*. 2000; 30(2): 137–142. [PubMed: 10832778]
- Hus V, Gotham K, Lord C. Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders*. 2014; 44(10):2400–2412. [PubMed: 23143131]
- Janvier YM, Harris JF, Coffield CN, et al. Screening for autism spectrum disorder in underserved communities: Early childcare providers as reporters. *Autism*. 2016; 20(3):364–373. [PubMed: 25991845]
- Johnson C, Myers S, Council on Children with Disabilities. Identification and Evaluation of Children With Autism Spectrum Disorders. *Pediatrics*. 2007; 120(5):1183–1215. [PubMed: 17967920]
- Khowaja M, Hazzard A, Robins D. Sociodemographic Barriers to Early Detection of Autism: Screening and Evaluation Using the M-CHAT, M-CHAT-R, and Follow-Up. *Journal of Autism and Developmental Disorders*. 2014; 45(6):1797–1808.
- King T, Tandon S, Macias M, et al. Implementing Developmental Screening and Referrals: Lessons Learned From a National Project. *Pediatrics*. 2010; 125(2):350–360. [PubMed: 20100754]
- Lord C, Luyster R, Gotham K. , et al. Autism Diagnostic Observation Schedule 2nd edition (ADOS-2) manual (Part II): Toddler module. Torrence: Western Psychological Services; 2012a.
- Lord C, Rutter M, DiLavore PC. , et al. Autism Diagnostic Observation Schedule-WPS Edition. Los Angeles: Western Psychological Services; 1999.
- Lord C, Rutter M, DiLavore PC. , et al. Autism Diagnostic Observation Schedule, 2nd edition (ADOS-2) manual (Part I): Modules 1–4. Torrence: Western Psychological Services; 2012b.
- Lord C, Rutter M, LeCouteur A. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*. 1994; 24:659–85. [PubMed: 7814313]

- MacDonald R, Parry-Cruwys D, Dupere S, et al. Assessing progress and outcome of early intensive behavioral intervention for toddlers with autism. *Research in Developmental Disabilities*. 2014; 35(12):3632–3644. [PubMed: 25241118]
- Meltzer S, Snyder J, Penrod J, et al. Gestational diabetes mellitus screening and diagnosis: a prospective randomised controlled trial comparing costs of one-step and two-step methods. *BJOG: An International Journal of Obstetrics & Gynaecology*. 2010; 117(4):407–415. [PubMed: 20105163]
- Mullen EM. *Mullen Scales of Early Learning: AGS Edition*. Vol. 1995. Circle Pines: American Guidance Service; 1995.
- Nicolaides K, Spencer K, Avgidou K, et al. Multicenter study of first-trimester screening for trisomy 21 in 75 821 pregnancies: results and estimation of the potential impact of individual risk-orientated two-stage first-trimester screening. *Ultrasound in Obstetrics and Gynecology*. 2005; 25(3):221–226. [PubMed: 15736186]
- Norris M, Lecavalier L. Screening Accuracy of Level 2 Autism Spectrum Disorder Rating Scales A Review of Selected Instruments. *Autism*. 2010; 14(4):263–284. [PubMed: 20591956]
- Orinstein AJ, Helt M, Troyb E, et al. Intervention for Optimal Outcome in Children and Adolescents with a History of Autism. *Journal of Developmental & Behavioral Pediatrics*. 2014; 35(4):247–256. [PubMed: 24799263]
- Paul R, Loomis R, Chawarska K. Adaptive Behavior in Toddlers Under Two with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*. 2014; 44(2):264–270. [PubMed: 21573835]
- Perry A, Cummings A, Geier J, et al. Predictors of outcome for children receiving intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders*. 2011; 5(1):592–603.
- Pierce K, Carter C, Weinfeld M, et al. Detecting, Studying, and Treating Autism Early: The One-Year Well-Baby Check-Up Approach. *The Journal of Pediatrics*. 2011; 159(3):458–465.e6. [PubMed: 21524759]
- Pinto-Martin J, Dunkle M, Earls M, et al. Developmental Stages of Developmental Screening: Steps to Implementation of a Successful Program. *American Journal of Public Health*. 2005; 95(11):1928–1932. [PubMed: 16195523]
- Radecki L, Sand-Loud N, O'Connor K, et al. Trends in the Use of Standardized Tools for Developmental Screening in Early Childhood: 2002–2009. *Pediatrics*. 2011; 128(1):14–19. [PubMed: 21708798]
- Ray-Subramanian C, Huai N, Ellis Weismer S. Brief Report: Adaptive Behavior and Cognitive Skills for Toddlers on the Autism Spectrum. *Journal of Autism and Developmental Disorders*. 2010; 41(5):679–684.
- Reichow B, Barton E, Boyd B, et al. Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews*. 2012
- Reichow B. Overview of Meta-Analyses on Early Intensive Behavioral Intervention for Young Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*. 2012; 42(4):512–520. [PubMed: 21404083]
- Reynolds C, Kamphaus R. *BASC-2: Behavior assessment system for children*. Second. Circle Pines: American Guidance Service; 2004.
- Robins DL, Fein D, Barton M. *The Modified Checklist for Autism in Toddlers, Revised, with Follow-up (M-CHAT-R/F)*. 2009. Self-published. www.mchatscreen.com
- Robins DL, Casagrande K, Barton M, et al. Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics*. 2014; 133(1):37–45. [PubMed: 24366990]
- Rotholz DA, Kinsman AM, Lacy KK, et al. Improving Early Identification and Intervention for Children at Risk for Autism Spectrum Disorder. *Pediatrics*. 2017; 139(2):e20161061. [PubMed: 28082407]
- Schopler E, Van Bourgondien M, Wellman G. , et al. *Childhood Autism Rating Scale-Second Edition*. Los Angeles: Western Psychological Association; 2010.
- Self T, Parham D, Rajagopalan J. Autism spectrum disorder early screening practices a survey of physicians. *Communication Disorders Quarterly*. 2015; 36(4):195–207.

- Shih S, Crowley S, Sheu J. Cost-effectiveness Analysis of a Two-stage Screening Intervention for Hepatocellular Carcinoma in Taiwan. *Journal of the Formosan Medical Association*. 2010; 109(1): 39–55. [PubMed: 20123585]
- Sparrow S, Cicchetti D, Balla D. *Vineland Adaptive Behavior Scales*. Second. Minneapolis: Pearson Assessments; 2005.
- Stone WL, Ousley OY. *STAT Manual: Screening Tool for Autism in Two-Year-Olds*. Vanderbilt University; 1997. Unpublished manuscript
- Stone WL, Coonrod E, Turner L, et al. Psychometric Properties of the STAT for Early Autism Screening. *Journal of Autism and Developmental Disorders*. 2004; 34(6):691–701. [PubMed: 15679188]
- Stone WL, McMahon C, Henderson L. Use of the Screening Tool for Autism in Two-Year-Olds (STAT) for children under 24 months: An exploratory study. *Autism*. 2008; 12(5):557–573. [PubMed: 18805947]
- Swinkels S, Dietz C, van Daalen E, et al. Screening for autistic spectrum in children aged 14 to 15 months. I: the development of the Early Screening of Autistic Traits Questionnaire (ESAT). *Journal of autism and developmental disorders*. 2006; 36(6):723–732. [PubMed: 16614790]
- Taylor L, Maybery M, Wray J, et al. Are there differences in the behavioural phenotypes of Autism Spectrum Disorder probands from simplex and multiplex families? *Research in Autism Spectrum Disorders*. 2015; 11:56–62.
- Zuckerman K, Mattox K, Donelan K, et al. Pediatrician Identification of Latino Children at Risk for Autism Spectrum Disorder. *Pediatrics*. 2013; 132(3):445–453. [PubMed: 23958770]

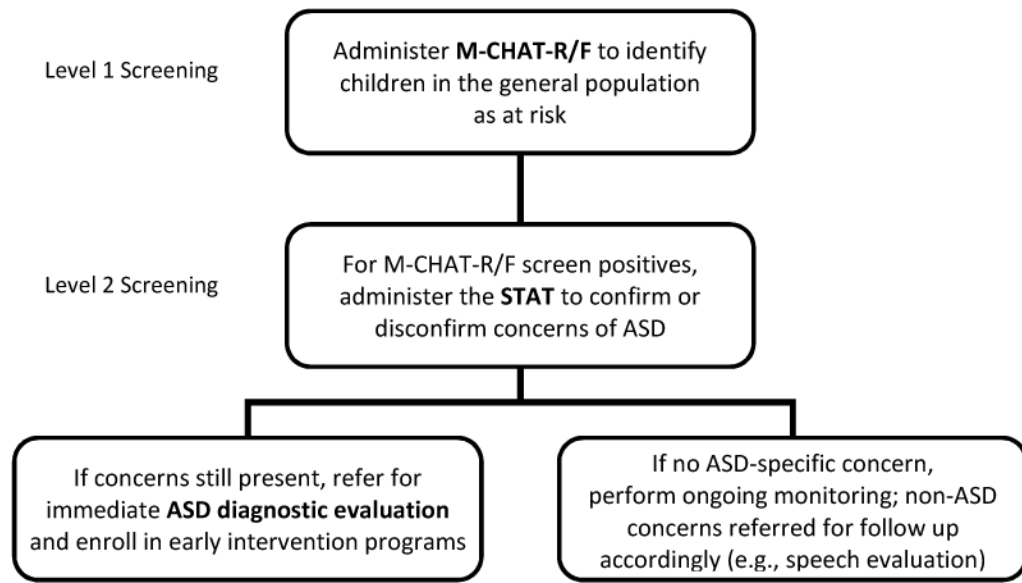


Figure 1.
Two-tiered approach to ASD screening.
M-CHAT-R/F: Modified Checklist for Autism in Toddlers, Revised, with Follow-up; STAT: Screening Tool for Autism in Toddlers and Young Children; ASD: autism spectrum disorder.

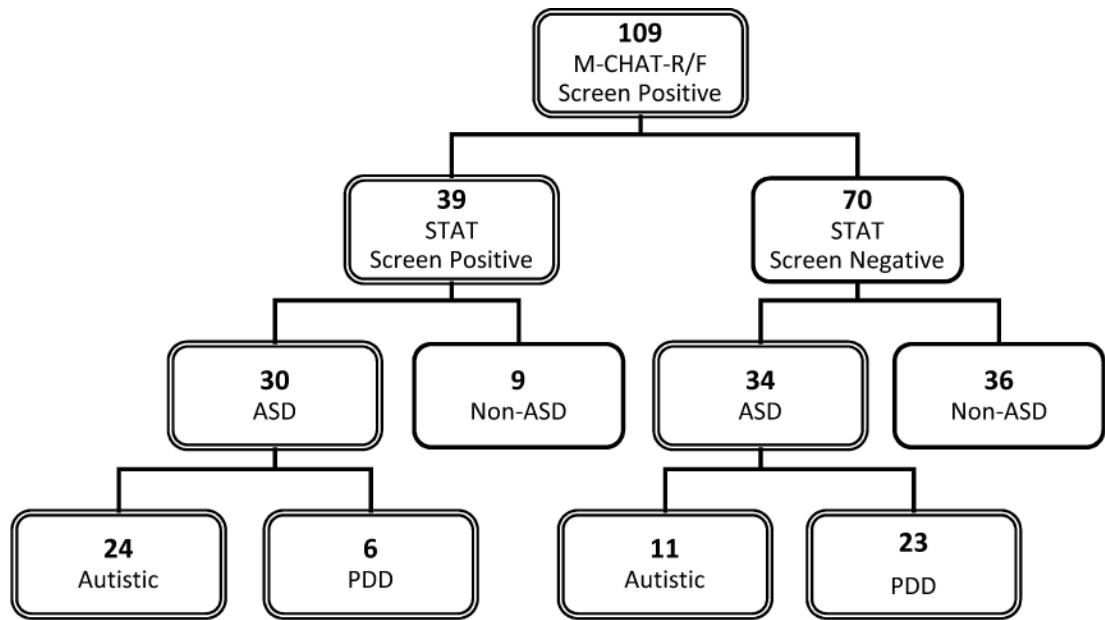


Figure 2. Two-tiered screening results with M-CHAT-R/F & STAT in high-risk sample. M-CHAT-R/F: Modified Checklist for Autism in Toddlers, Revised & Follow-Up; STAT: Screening Tool for Autism in Toddlers and Young Children; ASD: autism spectrum disorder; PDD: Pervasive Developmental Disorder, Not Otherwise Specified.

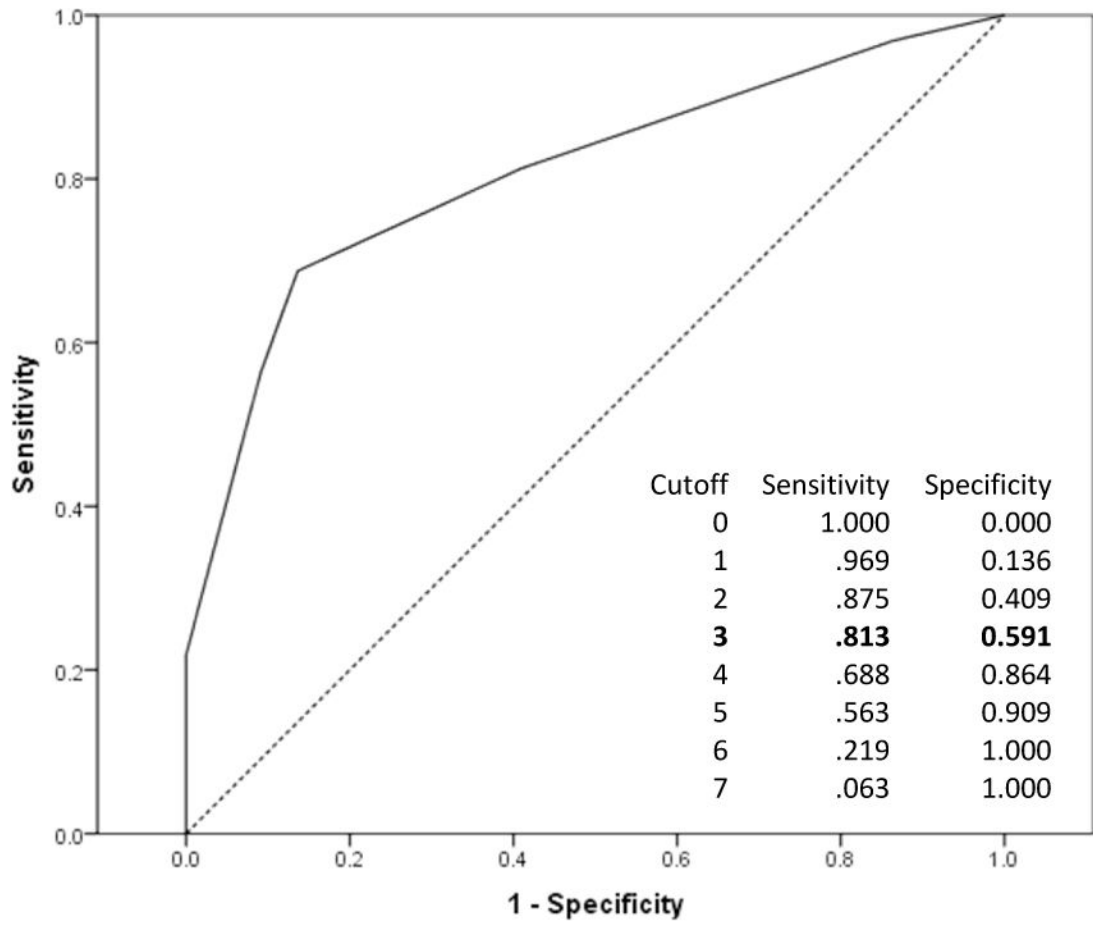


Figure 3.
ROC curve and psychometrics using new 7-item total score.

Table 1Sample Characteristics ($n = 109$)

	%	
Female	38.5	
African American	44.0	
Caucasian	37.6	
Diagnosis: Autism Spectrum Disorder	59.0	
Paper administration of M-CHAT-R	91.7	
	<i>M</i>	<i>SD</i>
Age at M-CHAT-R	20.34	3.38
Time from M-CHAT-R to Follow-Up ^a	2.69	1.95
Time from M-CHAT-R to STAT	4.48	2.71
Time from STAT to evaluation	.35	.63
Mullen Early Learning Composite	65.17	17.05
Vineland-II Adaptive Behavior Composite	78.81	10.68

M-CHAT-R: Modified Checklist for Autism in Toddlers, Revised; STAT: Screening Tool for Autism in Toddlers and Young Children.

^aSubsample that completed paper version of M-CHAT-R/F ($n=100$).

Table 2

Standardized Canonical Discriminant Function Coefficients

Item #	Item Description	Coefficient
7	Directing Attention: Bag of Toys	.539
6	Directing Attention: Puppet	.366
1	Play: Turn-Taking	.363
2	Play: Doll	.321
4	Requesting: Food	.314
3	Requesting: Bubbles	-.271
9	Imitation: Shake Rattle	.270
8	Directing Attention: Noisemaker	-.178
10	Imitation: Roll Car	-.170
5	Directing Attention: Balloon	.140
12	Imitation: Hop Dog	-.117
11	Imitation: Drum Hands	.034

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Correlations between both STAT scores and severity of autism symptoms, cognitive abilities, and adaptive functioning

	STAT Total (0–4)	STAT DFA 7-item Total (0–7)
ADOS Total	.51*	.55*
ADOS Social Affect	.55*	.58*
ADOS Restricted & Repetitive Behaviors	.34*	.39*
Mullen Early Learning Composite	-.46*	-.43*
Vineland-II Adaptive Behavior Composite	-.46*	-.44*

STAT: Screening Tool for Autism in Toddlers and Young Children; DFA: Discriminant Function Analysis;

ADOS: Autism Diagnostic Observation Schedule.

* $p < .001$.

Table 4

Clinical data across subsamples with ASD and non-ASD outcomes

	ASD		Non-ASD	
	STAT Positive (<i>n</i> =30) <i>M</i> (SD)	STAT Negative (<i>n</i> =34) <i>M</i> (SD)	STAT Positive (<i>n</i> =9) <i>M</i> (SD)	STAT Negative (<i>n</i> =36) <i>M</i> (SD)
ADOS Total Severity	7.33 ^a (1.75)	5.85 ^b (2.19)	2.67 (1.50)	1.94 (1.35)
ADOS SA Severity	7.50 ^a (1.76)	6.06 ^b (1.74)	3.22 (1.56)	2.28 (1.34)
ADOS RRB Severity	7.37 (1.99)	5.88 (2.47)	4.33 (1.94)	2.97 (2.43)
Mullen ELC	54.90 ^a (8.37)	68.59 ^b (19.36)	59.00 (15.91)	72.06 (16.35)
Vineland-II ABC	71.57 ^a (9.48)	80.21 ^b (10.22)	79.22 (7.45)	83.54 (9.82)

ASD: autism spectrum disorder; STAT: Screening Tool for Autism in Toddlers and Young Children; ADOS: Autism Diagnostic Observation Schedule; SA: Social Affect; RRB: Restricted and Repetitive Behaviors; ELC: Early Learning Composite; ABC: Adaptive Behavior Composite.

Note. Data compare results based on STAT performance within each diagnostic group utilizing independent samples *t*-tests with Sidak-Bonferroni correction ($\alpha=.010$).

a,b different letters indicate significant differences between subsamples