Title

Utility of the Child Behavior Checklist as a Screener for Autism Spectrum Disorder

Permalink

https://escholarship.org/uc/item/3js4v32s

Journal

Autism Research, 9(1)

ISSN

1939-3792

Authors

Havdahl, K Alexandra
von Tetzchner, Stephen
Huerta, Marisela
et al.

Publication Date

2016

DOI

10.1002/aur.1515

Peer reviewed

Utility of the Child Behavior Checklist as a Screener for Autism Spectrum Disorder

K. Alexandra Havdahl, **Stephen von Tetzchner**, **Marisela Huerta**, **Catherine Lord**, and **Somer L. Bishop**

Center for Autism and the Developing Brain, Weill Cornell Medical College, White Plains, New York (K.A.H., M.H., C.L.); Lovisenberg Diaconal Hospital, Oslo, Norway (K.A.H.); Norwegian Institute of Public Health, Oslo, Norway (K.A.H.); Department of Psychology, University of Oslo, Oslo, Norway (S.T.); Department of Psychiatry, University of California San Francisco, California (S.L.B.)

Abstract

The Child Behavior Checklist (CBCL) has been proposed for screening of autism spectrum disorders (ASD) in clinical settings. Given the already widespread use of the CBCL, this could have great implications for clinical practice. This study examined the utility of CBCL profiles in differentiating children with ASD from children with other clinical disorders. Participants were 226 children with ASD and 163 children with attention-deficit/hyperactivity disorder, intellectual disability, language disorders, or emotional disorders, aged 2–13 years. Diagnosis was based on comprehensive clinical evaluation including well-validated diagnostic instruments for ASD and cognitive testing. Discriminative validity of CBCL profiles proposed for ASD screening was examined with area under the curve (AUC) scores, sensitivity, and specificity. The CBCL profiles showed low discriminative accuracy for ASD (AUC 0.59–0.70). Meeting cutoffs proposed for ASD was associated with general emotional/behavioral problems (EBP; mood problems/ aggressive behavior), both in children with and without ASD. Cutoff adjustment depending on EBP-level was associated with improved discriminative accuracy for school-age children. However, the rate of false positives remained high in children with clinical levels of EBP. The results indicate that use of the CBCL profiles for ASD-specific screening would likely result in a large number of misclassifications. Although taking EBP-level into account was associated with improved discriminative accuracy for ASD, acceptable specificity could only be achieved for school-age children with below clinical levels of EBP. Further research should explore the potential of using the EBP adjustment strategy to improve the screening efficiency of other more ASD-specific instruments.

Keywords

early detection; diagnosis; emotional/behavioral problems; Child Behavior Checklist (CBCL)

Address for correspondence and reprints: Karoline Alexandra Havdahl, Weill Cornell Medical College, Center for Autism and the Developing Brain, 21 Bloomingdale Road, White Plains, NY, 10605. kah2028@med.cornell.edu or alha@fhi.no.

Introduction

Diagnosis of ASD is often difficult due to the heterogeneity in severity and constellations of ASD symptoms, variation in symptom presentation with developmental level and age, and common co-occurrence of other psychiatric conditions. Differential diagnosis is further complicated by the fact that social difficulties and repetitive behaviors are also seen in children with non-ASD diagnoses such as attention-deficit/hyperactivity disorder (ADHD), language disorders, intellectual disability (ID), and emotional disorders. Although well-validated diagnostic instruments are available to aid in differential diagnosis, in-depth assessment of ASD is time intensive and requires clinicians with extensive training and experience with ASD [Huerta & Lord, 2012]. This has resulted in an increasing effort to establish reliable, valid, and cost-efficient instruments that can support clinicians in determining a need for further ASD evaluation.

The Child Behavior Checklist (CBCL) is a well established and widely used parent-completed measure of emotional, behavioral, and social problems in children aged 1.5–5 years and 6–18 years [Achenbach & Rescorla, 2000, 2001]. It was developed to assess a range of problem behaviors rather than ASD in particular, and discriminates well between clinic-referred and non-referred children [Achenbach & Rescorla, 2000, 2001]. Recently, the instrument developer proposed that the CBCL is also useful for ASD-specific screening within clinical settings [Achenbach & Rescorla, 2013].

Multiple CBCL scales and profiles have been suggested for ASD screening. The CBCL/1.5–5 Withdrawn and Pervasive Developmental Problems (PDP) scales have been reported to have high accuracy in distinguishing pre-schoolers with ASD from preschoolers with other disorders (AUC 0.85–0.94) [Muratori et al., 2011; Narzisi et al., 2013]. The CBCL/6–18 scales Withdrawn/Depressed, Social problems, and Thought problems have also been found to differentiate well between school-age children with ASD and non-ASD disorders [Biederman et al., 2010; Duarte, Bordin, de Oliveira, & Bird, 2003; Ooi, Rescorla, Ang, Woo, & Fung, 2011]. However, generalizability of these findings is potentially limited by methodological issues, such as exclusion of children with relevant differential diagnoses (e.g., ADHD, language/cognitive impairments). Additionally, validity of the results may be limited by sampling children with ASD with a high degree of general behavior problems, especially when comparison is made with a non-ASD group with lower levels of behavior problems. Other studies suggest that emotional/behavior problems (EBP), such as aggressive behavior and mood problems, contribute substantially to elevated scores on other ASD screening tools [Charman et al., 2007; Hus, Bishop, Gotham, Huerta, & Lord, 2013]. Hus et al. [2013] suggested that taking non-ASD-specific behavior problems into account may be needed to appropriately interpret scores on ASD screeners.

Some studies have found lower accuracy of CBCL profiles in identifying children with ASD in the context of children with other clinical problems (AUC 0.67–0.75) [Myers, Gross, & McReynolds, 2014; Ooi et al., 2014; Rescorla, Kim, & Oh, 2014; So et al., 2013]. Comparison of results across studies is difficult due to variation in ascertainment methods and limited sample characterization in terms of autism symptom severity, intellectual ability, and language level. To our knowledge, no previous studies on this topic have completed in-

depth assessment of ASD for children included in both the ASD group and the comparison group. Additionally, there is little information about the effect of including children with previous ASD diagnoses, or about child characteristics found to influence other ASD screening tools, such as intellectual ability, age, and gender [Charman et al., 2007; Cholemkery, Mojica, Rohrmann, Gensthaler, & Freitag, 2014; Corsello et al., 2007].

The variability in discriminative accuracy across studies clearly warrant further examination of the CBCL's validity in distinguishing children with ASD from children with other clinical disorders. Screening misclassification may lead to inappropriate clinical decisions in the assessment process and/or loss of valuable time for appropriate interventions [Norris & Lecavalier, 2010]. This study examines the utility of CBCL scales proposed for ASD-specific screening to distinguish children with ASD from children with non-ASD disorders commonly seen in ASD diagnostic clinics. The study also explores factors that may help explain the variability in results.

Methods

Participants

The sample consisted of 407 children aged 2–13 years who had been assessed for ASD as part of a research study of autism symptoms in children with non-ASD diagnoses of ADHD, ID, language disorders, or emotional disorders. For the current study, the only exclusionary criteria were incomplete CBCL data (>8 missing items, n = 8), and having no DSM-IV-TR disorder (n = 10). The participants were recruited mainly through clinic intake/referral and flyers/website communication, either in the Division of Developmental and Behavioral Pediatrics at Cincinnati Children's Hospital Medical Center (CCHMC), or at the University of Michigan Autism and Communications Disorders Center (UMACC).

The majority of the parents had some college (n = 143) or a higher education level (n = 166), and fewer had H.S. diploma without college (n = 37) or less (n = 10, missing n = 33). No significant difference was found between sites in parent education level, $X^2 = 2.87$, df = 3, P = 0.41. The proportion of children of non-white/Caucasian ethnicity was higher at CCHMC (n = 66, missing n = 1) compared to at UMACC (n = 58, missing n = 1), $X^2 = 9.68$, df = 1, P <0.01.

The majority of the children without ASD came from CCHMC (88%). Given that these children had previously received non-ASD diagnoses and were not referred for ASD concern, they are likely representative of children presenting to general developmental disabilities/psychiatric clinics for assessment. The proportions of individual non-ASD diagnoses are presented in Table 1. The children with ASD (DSM-IV-TR: autistic disorder, n = 156, pervasive developmental disorder-not otherwise specified, n = 65, Asperger syndrome, n = 5) came mainly from UMACC (80%). Nearly half received the ASD diagnosis for the first time through the research study (preschoolers: 50%, school-aged: 43%).

Measures and Procedure

This research was approved by the Institutional Review Board at CCHMC and UMACC. Prior to participation, all caregivers signed an informed consent form.

Parents completed the CBCL prior to the diagnostic evaluation, with a mean time lag of 15 days (SD = 35). The individual CBCL scales are presented in Table 2. All children underwent a comprehensive clinical evaluation, including well-validated diagnostic instruments for ASD [i.e., the Autism Diagnostic Interview-Revised, ADI-R; Rutter, LeCouteur, & Lord, 2003 and the Autism Diagnostic Observation Schedule, ADOS; Lord, Rutter, DiLavore, & Risi, 1999; Lord et al., 2012], the Vineland adaptive behavior scales-II [Sparrow, Cicchetti, & Balla, 2005], and cognitive testing: the Differential Ability Scales-II [Elliott, 1990; n = 330] or the Mullen Scales of Early Learning, [Mullen, 1995; n = 58]. The assessment also included measures relevant for establishing non-ASD diagnoses, such as the Conners' Parent Rating Scale-Revised [Conners, Sitarenios, Parker, & Epstein, 1998], the Spence Children's Anxiety Scale [Spence, 1998], and the Multidimensional Anxiety Scale for Children [March, Parker, Sullivan, Stallings, & Conners, 1997]. Following completion of all measures, clinicians met to discuss their impressions and assign a consensus diagnosis. Although the CBCL was available at time of diagnosis, this instrument was not used in determining the presence or absence of ASD.

Data Analysis

Analyses were carried out separately for the CBCL/1.5–5 and the CBCL/6–18, using the Statistical Package for Social Sciences (SPSS) version 21. Significance level was set at alpha = 0.05 (two-tailed). Characteristics of the ASD and non-ASD groups were compared using chi square tests (Fisher's exact test if cells <5 observations) and t-tests.

First, we examined whether the CBCL scales suggested for ASD screening (i.e., Withdrawn, PDP, Withdrawn/depressed, Social problems, and Thought problems) showed diagnostic group differences when controlling for other child characteristics. Multivariate Analysis of Covariance (MANCOVA) was used to examine diagnostic group differences on (a) composite scales, (b) syndrome scales, and (c) DSM-oriented scales, with gender, nonverbal IQ, and age as covariates. Raw scores were used in the MANCOVA, as recommended by Achenbach and Rescorla [2000, 2001]. Individual ANCOVAs were only analyzed if the MANCOVA was significant. Effect sizes are reported as partial eta squared (η_p^2), interpreted as small: 0.01–0.05, medium: 0.06–0.13, and large: 0.14.

Logistic regression was used to determine whether scale combinations resulted in incremental discriminative validity compared with the individual scales. Discriminative validity was examined using area under the curve (AUC) scores from nonparametric receiver operating curve (ROC) analyses, which is a plot of true positive vs. false positive results. Swets [1988] suggested the following benchmarks for interpreting AUC scores: 0.50–0.70 (low accuracy), 0.70–0.90 (moderate accuracy), and >0.90 (high accuracy). A sample size calculation, using the StatsToDo website (https://www.statstodo.com/SSizSenSpc_Pgm.php), indicated that 50 cases in each group were needed to detect a difference between chance-level and moderate discrimination (AUC = 0.50/0.70, α = 0.05,

power = 0.80). For the profile demonstrating the highest AUC-score in each age group, we calculated sensitivity, specificity, and positive likelihood ratio (LR+). Confidence intervals (95%) were calculated based on the Wilson score method [Newcombe, 1998]. T scores were used to facilitate comparison with previous studies.

Stratified analyses were performed to examine whether discriminative accuracy was associated with level of EBP, ID, and/or previous ASD diagnosis. The CBCL has multiple scales intended to capture emotional problems (e.g., Internalizing, Emotionally reactive, Anxious/depressed, Anxiety problems, and Affective problems) and behavioral problems (e.g., Externalizing, Attention problems, Attention deficit/hyperactivity problems, Oppositional/defiant problems). In operationalizing clinically significant level of EBP, avoiding overlap with core ASD behaviors was a priority. Therefore, scales with item content clearly overlapping with core ASD behaviors were not considered (e.g., Emotionally reactive, Internalizing). Few studies have examined concordance between CBCL scales and co-occurring emotional/behavioral disorders in children with ASD. An exception is a recent study of school-aged children with ASD, finding the highest discriminative validity for the Affective problems and Aggressive behavior scales (AUC = 0.90) [Gjevik, Sandstad, Andreassen, Myhre, & Sponheim, 2015]. To avoid the multiple comparisons problem, we based the choice of the particular emotional and behavioral scale on this finding. Therefore, EBP-level was operationalized as high when T score (age- and gender-normed) on Aggressive behavior and/or Affective problems was in the clinical range (70). For the EBP classification to be useful in children with problems specific to the emotional or behavioral domain, high EBP was defined as scoring in the clinical range on either of the scales (results were very similar when using only one of the scales).

All results should be interpreted in light of their confidence intervals. Charman et al. [2007] found a difference in specificity of 0.41 and 0.93 for another ASD screener between subgroups with high and low EBP. A sample size calculation indicated that 13 cases in each group were needed to have 80% power to detect a difference of this size ($\alpha = 0.05$; StatsToDo).

Results

Sample Characteristics

As shown in Table 1, there were large differences in ADOS scores between the ASD and non-ASD groups. The ASD group also showed lower intellectual ability, with significant differences in verbal IQ in both age samples, and in nonverbal IQ in the preschool sample. No significant differences were found for age or gender proportions. Among children with non-ASD disorders, the proportion with language disorders was higher in the preschoolers, whereas the proportion with ADHD and emotional disorders was higher in the school-age children. The prevalence of high EBP was 33% in the total sample, with no significant differences between the ASD and non-ASD groups. The two scales comprising EBP-level did not significantly correlate with age, nonverbal IQ, or verbal IQ (Pearson's r ranged from −0.09 to 0.09, P 0.16).

Group Differences on the CBCL

Table 2 presents mean raw CBCL scores and MANCOVA results for the ASD and non-ASD groups (mean T scores are provided as supplementary information). Controlling for gender, age, and nonverbal IQ, preschoolers with ASD scored significantly higher than preschoolers with non-ASD disorders on Withdrawn and PDP (medium effect sizes, ES). The ASD group also scored significantly higher on Total problems, Internalizing, Emotionally reactive, Aggressive behavior, and Anxiety problems (small ES). In the school-age sample, the ASD group scored significantly higher than the non-ASD group only on the scales suggested for ASD screening (i.e., Withdrawn/depressed, Social problems, and Thought problems, small-to-medium ES), controlling for gender, age, and nonverbal IQ.

Overall Discriminative Validity

As shown in Table 3, overall discriminative validity of the two CBCL/1.5–5 scales proposed for ASD screening was in the low range (AUC 0.68–0.69). Logistic regression showed no incremental discriminative value of combining the scales. Only Withdrawn made a significant unique contribution to discrimination (B = 0.22, P = 0.01), while the nonoverlapping items from PDP did not contribute significantly (B = 0.00, P = 0.99), $\chi^2(2) = 15.02$, P <0.01. Due to similar findings, further results are only presented for Withdrawn.

The CBCL/6–18 scales suggested for ASD screening also resulted in AUC-scores in the low range (AUC = 0.59–0.67). Logistic regression showed that combining the scales had incremental discriminative value compared to the individual scales. Withdrawn/depressed and Thought problems made statistically significant unique contributions to discrimination (B = 0.06, P <0.01 and B = 0.05, P <0.01, respectively), whereas Social problems did not contribute significantly (B = –0.02, P = 0.32), $\chi^2(3) = 29.04$, P <0.01. The aggregated scale of T scores from Withdrawn/depressed and Thought problems, hereafter referred to as Withdrawn-Thought Problems (WTP), yielded an AUC-score of 0.70.

Given the site differences between the ASD and non-ASD groups, we examined the possible covariate effect of site (UMACC vs. CCHMC) using ROC regression in Stata version 13. Site did not show a significant covariate effect on either the preschool Withdrawn scale (P = 0.87) or the school-age WTP scale (P = 0.85).

Sensitivity, Specificity, and Likelihood Ratio

Sensitivity, specificity, and LR+ of the Withdrawn and WTP scales was examined at two previously suggested T score cutoffs of 65 and 62 [Muratori et al., 2011; Narzisi et al., 2013], using the aggregated mean scale cutoff when combining scales (130 and 124 for WTP) [Biederman et al., 2010]. At the higher cutoff consistent with the CBCL "borderline clinical" cut-point, sensitivity and specificity was 63% (95% CI = 53–73) and 65% (95% CI = 51–77) for Withdrawn, and 58% (95% CI = 50–68) and 68% (95% CI = 58–76) for WTP, respectively. LR+ was 1.8 for both Withdrawn (95% CI = 1.2–2.7) and WTP (95% CI = 1.3–2.5).

The lower cutoff resulted in moderate sensitivity (74% for Withdrawn, 78% for WTP) and low specificity (53% for Withdrawn, 55% for WTP). Change in probability of ASD

diagnosis given scores above the lower cutoff was small both for Withdrawn (1.6) and WTP (1.7). The cutoff required to identify at least 80% of children with ASD resulted in specificity of 39% for Withdrawn (95% CI = 26–51, cutoff 58) and 53% for WTP (95% CI = 43–63, cutoff 123).

Factors Associated With Discriminative Validity

Table 4 presents the results of the subgroup analyses for the more sensitive lower cutoff by level of EBP, ID, and previously/first diagnosed ASD. Subgroup analysis by gender was attempted, but was not possible due to confounding of gender and high EBP within children with ASD, with significantly higher proportion of EBP in girls compared to boys in preschoolers (53% vs. 29%), $\chi^2(1, N = 104) = 4.20$, P = 0.04, and school-age children (55% vs. 31%), $\chi^2(1, N = 122) = 5.75$, P = 0.02). There was no significant difference in the proportions of high EBP between girls and boys with non-ASD disorders in preschoolers (14% vs. 23%), Fisher's exact P = 0.71, or in school-aged children (32% vs. 33%), $\chi^2(1, N = 106) = 0.01$, P = 0.93).

Level of EBP

Discriminative utility of the Withdrawn and WTP showed substantial variability depending on EBP-level. For both scales, discriminative validity was in the moderate range for children with low EBP (AUC = 0.70–0.79) and in the low range for children with high EBP (AUC = 0.62).

With regard to the CBCL/6–18 WTP, scores at or above 124 were associated with a 3.2 increase in likelihood of ASD among children with low EBP, in contrast to no increase among children with high EBP (1.0). Optimal cutoffs (maximized specificity with sensitivity 80%) were widely differing in children with high compared to low EBP-level. In the low EBP subgroup, a cutoff of 117 correctly classified 82% (95% CI = 71–89) of children with ASD and 62% (95% CI = 50–73) of children with non-ASD disorders. For children with high EBP, compared to cutoff 124, a cutoff of 134 resulted in improved specificity from 6% (95% CI = 1–13) to 40% (95% CI = 24–58) while maintaining sensitivity at 81% (95% CI = 67–91) (see Fig. 1).

Although a similar pattern was found for the CBCL/1.5–5 Withdrawn, CIs were wider, especially in the small high EBP subgroup (n = 46). In the larger low EBP subgroup (n = 115), discriminative accuracy was somewhat lower than for the school-age low EBP subgroup (AUC 0.70 vs. 0.79). The cutoff required to identify at least 80% of preschoolers with ASD in the low EBP subgroup, resulted in only 33% specificity (cutoff 54, sensitivity: 87%). Thus, it was not possible to achieve acceptable discriminative accuracy by using adjusted cutoffs.

Intellectual Disability

Due to few children with ID in the preschool non-ASD group (n = 5), this analysis was only performed for the school-age sample. Although discriminative accuracy of the WTP was in the moderate range for children without ID (AUC = 0.73) and in the low range for children with ID (AUC = 0.59), the CIs were highly overlapping.

Previously Diagnosed ASD

Limiting the preschool ASD group to previously diagnosed vs. children diagnosed for the first time, discriminative accuracy of the Withdrawn scale was in the moderate (AUC = 0.74) and low range (AUC = 0.64), respectively. Sensitivity of the lower cutoff was within acceptable limits (80%) only for preschoolers with previous ASD diagnoses. However, the CIs of the estimates overlapped.

WTP differentiated school-age children with and without ASD similarly when the ASD group was limited to children previously diagnosed (AUC = 0.70) as to children first diagnosed (AUC = 0.71).

Discussion

Children with ASD scored significantly higher than children with non-ASD disorders on CBCL scales proposed for ASD screening (i.e., Withdrawn, PDP, Withdrawn/depressed, Social problems, and Thought problems), when controlling for other child characteristics. The CBCL/1.5–5 scales Withdrawn and PDP showed similar differentiation, whereas a combination of the CBCL/6–18 scales Withdrawn/depressed and Thought problems differentiated best. However, the scales showed low discriminative validity when used to distinguish between individual children with ASD and non-ASD disorders (AUC 0.59–0.70). Scores above previously suggested cutoffs were associated with only a small increase in probability of ASD diagnosis (all 1.8).

There is an inherent tradeoff between maximizing sensitivity and minimizing false positives, and priority depends on the purpose of the instrument. Considering that the CBCL has been proposed for screening rather than diagnosis, sensitivity may be considered the highest priority. The cutoff required to identify at least 80% of children with ASD in this study was lower than found in previous studies. Compared to reported sensitivity of 78–90% [Biederman et al., 2010; Myers et al., 2014; Narzisi et al., 2013], sensitivity in this study was 58–63% at the threshold consistent with the CBCL "borderline clinical" problems cutoff (65 for individual narrow-band scales; average scale score for scale combinations). Limited sample characterization in previous studies makes comparison difficult, which is problematic given that sample characteristics influence our ability to predict screening efficiency in the intended population. Biederman et al. [2010] reported higher sensitivity in their ASD sample characterized by a high level of general behavior problems, consistent with the subgroup showing the highest sensitivity in this study. In some studies, lack of representation of children with milder ASD presentations (i.e., DSM-IV/ICD diagnoses other than autistic disorder) is likely to have contributed to higher sensitivity estimates [Myers et al., 2014; Ooi et al., 2011].

Utility of the CBCL to identify children in need of further ASD assessment requires specificity within acceptable limits with regard to resources needed to resolve false positive cases and potential loss of time for appropriate interventions. At the threshold necessary to identify at least 80% of children with ASD, specificity was low (39–53%). This is consistent with low-to-moderate specificity found for CBCL profiles in two other studies that included a range of non-ASD disorders [Myers et al., 2014; So et al., 2013]. The results indicate that

the CBCL scales would likely result in a large number of false positives if used to screen for ASD in clinical settings. False positive screening could lead to a narrowing of assessment focus, possibly at the expense of more appropriate alternatives. Resolving false positive cases can cost valuable time and resources and/or delay delivery of appropriate interventions. Additionally, unwarranted referrals to ASD specialty clinics could give rise to needless emotional distress and economic expenses for families [Sikora, Hall, Hartley, Gerrard-Morris, & Cagle, 2008].

In line with previous findings for other ASD screening tools [Charman et al., 2007], specificity was especially low in children with high EBP, with 74–92% of children with non-ASD disorders misclassified when using proposed cutoffs. Although statistical control is not available in clinical practice, clinicians may nevertheless need to take into account the level of EBP when interpreting ASD screening results. In this study, the age and gender normed CBCL scales Affective problems and Aggressive behavior were used to define an easily applicable indicator of high EBP (either scale 70). The optimal cutoff maximizing specificity with high sensitivity (80%) differed widely between the subgroups stratified by EBP-level. For the WTP, use of EBP-level specific cutoffs resulted in greatly improved specificity in children with high EBP, while maintaining sensitivity above 80% in both EBP subgroups and with 62% specificity in children with low EBP. Although this strategy led to substantially improved discriminative accuracy, the rate of false positives was still high in children with high EBP.

Although EBP-level also seemed to moderate the discriminative accuracy of the preschool Withdrawn scale (i.e., differing likelihood ratios by EBP-level), it was not possible to achieve similar overall improvement of discriminative accuracy with the use of EBP-level-specific cutoffs. Unlike the school-age WTP scale, the preschool Withdrawn scale showed poor discriminative accuracy among children without clinically significant EBP. There may be several explanations for the variability in discriminative accuracy between the preschoolers and school-aged children with low EBP. First, different scales were used for the two age groups with only the school-age WTP including items related to the repetitive behavior symptom domain (e.g., "Repeats certain acts over and over," "Can't get his/her mind off certain thoughts"). The lack of representation of this core ASD symptom domain could help explain the poor discriminative accuracy even in the low EBP subgroup. Another contributing factor could be the relatively higher proportion of children with language disorders in the preschool compared to the school-aged non-ASD group. Withdrawn has been found to be the most commonly elevated narrow-band CBCL scale in pre-schoolers with language disorders [Maggio et al., 2014]. Third, given that the choice of the particular EBP scales was based on research with school-aged children [Gjevik et al., 2015], alternative EBP classifications could potentially be more useful for defining adjusted cutoffs on ASD screening scales in preschoolers. Finally, because problem behaviors and symptoms may be less differentiated in very young children, it is possible that adjusting for EBP-level has less utility for improving discriminative accuracy of ASD screeners in preschoolers compared to older children. Future studies could examine this with the use of the same ASD screener across age groups.

The results of this study demonstrate the importance of taking moderating factors such as EBP-level into account when evaluating the discriminative validity of screening tools for ASD. Given that high EBP was associated with increased likelihood of meeting cutoffs on the scales proposed for ASD screening, estimates of discriminative accuracy could vary according to the distribution of EBP-level in a particular sample [see Janes & Pepe, 2008]. Depending on whether the rate of EBP is higher in the ASD or in the non-ASD comparison group, overall estimates of discriminative accuracy could be overestimated or underestimated, respectively. Although there was no significant difference on the EBP-level classifier between the diagnostic groups in this study, the preschoolers with ASD had somewhat higher scores on several emotional/behavioral scales including Aggressive behavior compared to preschoolers without ASD. Thus, given that higher EBP would be expected to be associated with higher likelihood of meeting cutoffs on the Withdrawn scale, the poor overall discriminative accuracy in the preschool age group is perhaps even more concerning.

This study adds to the literature on this topic in several ways, including (a) use of a well-characterized sample of children with ASD and children with previous diagnoses of non-ASD disorders who all completed a comprehensive diagnostic evaluation of ASD; (b) exploring factors that may help explain the variability in results across studies; and (c) presenting a strategy for taking EBP-level into account to improve the discriminative accuracy for ASD. The results must also be interpreted in light of some methodological limitations. The sample consisted mainly of children with relatively high intellectual ability, and our findings may not generalize to more cognitively impaired children. However, given that the CBCL has not been normed for children with ID, the sample may be especially relevant to the population for which it is intended. In common with Biederman et al. [2010], the sample included children with previous diagnoses, and child behavior rating may differ depending on whether parents are aware of the presence of a diagnosis. However, if previous diagnosis leads to more parent awareness of behaviors associated with the particular diagnosis, this should have contributed to higher discriminative accuracy for ASD rather than lower, further supporting our finding of low overall accuracy. Notably, due to small subgroups reflected in wide confidence intervals, power was limited and replication with larger samples is needed to yield precise estimates. Another limitation is that the screening scales and the EBP classification were derived from the same instrument. Finally, given that there is little knowledge about which CBCL scales are most accurate in capturing EBP in children with and without ASD and at different age levels, future studies should examine which scales and cutoffs are most useful for determining EBP-levels.

Due to the widespread use of the CBCL in clinical settings worldwide, reports of its utility for ASD-specific screening could have substantial implications for practice. Although the CBCL is useful in providing information about a range of behavioral functions and for identifying children with behavior problems (first level screening), the results of this study suggest that the CBCL scales are not useful for ASD-specific screening. Although adjustment for EBP-level improved specificity, it was not possible to achieve acceptable levels of sensitivity and specificity due to moderate discriminative validity even within subgroups of children with low EBP, without ID, and with previous ASD diagnoses. However, the strategy of using the CBCL to define EBP-level and applying EBP-level

specific cutoffs could potentially improve the screening efficiency of other tools that are more ASD-specific, such as the Social Responsiveness Scale (Constantino & Gruber, 2005) or the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003). It may also be possible to improve the discriminative validity of diagnostic instruments for ASD, such as the ADI-R and the ADOS, by taking EBP-level into account. Future research is needed to examine this.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Achenbach, TM.; Rescorla, LA. Manual for the ASEBA preschool forms & profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2000.

Achenbach, TM.; Rescorla, LA. Manual for the ASEBA school-age forms & profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2001.

Achenbach, TM.; Rescorla, LA. Achenbach system of empirically based assessment. In: Volkmar, F., editor. Encyclopedia of autism spectrum disorders. New York, NY: Springer; 2013. p. 31-39.

Biederman J, Petty CR, Fried R, Wozniak J, Micco JA, Henin A, Faraone SV. Child Behavior Checklist clinical scales discriminate referred youth with autism spectrum disorder: A preliminary study. Journal of Developmental & Behavioral Pediatrics. 2010; 31(6):485–490. [PubMed: 20585266]

Charman T, Baird G, Simonoff E, Loucas T, Chandler S, Meldrum D, Pickles A. Efficacy of three screening instruments in the identification of autistic-spectrum disorders. The British Journal of Psychiatry. 2007; 191(6):554–559. [PubMed: 18055961]

Cholemkery H, Mojica L, Rohrmann S, Gensthaler A, Freitag CM. Can autism spectrum disorders and social anxiety disorders be differentiated by the Social Responsiveness Scale in children and adolescents? Journal of Autism and Developmental Disorders. 2014; 44(5):1168–1182. [PubMed: 24682652]

Conners CK, Sitarenios G, Parker JDA, Epstein JN. The revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. Journal of Abnormal Child Psychology. 1998; 26(4):257–268. [PubMed: 9700518]

Constantino, JN.; Gruber, C. The social responsiveness scale. Los Angeles, CA: Western Psychological Services; 2005.

Corsello C, Hus V, Pickles A, Risi S, Cook EH, Leventhal BL, Lord C. Between a ROC and a hard place: Decision making and making decisions about using the SCQ. Journal of Child Psychology and Psychiatry. 2007; 48(9):932–940. [PubMed: 17714378]

Duarte CS, Bordin IA, de Oliveira A, Bird H. The CBCL and the identification of children with autism and related conditions in Brazil: Pilot findings. Journal of Autism and Developmental Disorders. 2003; 33(6):703–707. [PubMed: 14714937]

Elliott, C. Differential ability scales. 2. San Antonio, TX: Harcourt Assessment; 2007.

Gjevik E, Sandstad B, Andreassen OA, Myhre AM, Sponheim E. Exploring the agreement between questionnaire information and DSM-IV diagnoses of comorbid psychopathology in children with autism spectrum disorders. Autism. 2015; 19(4):433–442. [PubMed: 24637430]

Huerta M, Lord C. Diagnostic evaluation of autism spectrum disorders. Pediatric Clinics of North America. 2012; 59(1):103–111. [PubMed: 22284796]

Hus V, Bishop S, Gotham K, Huerta M, Lord C. Factors influencing scores on the Social Responsiveness Scale. Journal of Child Psychology and Psychiatry. 2013; 54(2):216–224. [PubMed: 22823182]

Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: An old concept in a new setting. American Journal of Epidemiology. 2008; 168(1):89–97. [PubMed: 18477651]

Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S.; Gotham, K.; Bishop, SL. Autism Diagnostic Observation Schedule–2nd edition (ADOS-2). Torrance, CA: Western Psychological Services; 2012.

Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S. Autism Diagnostic Observation Schedule (ADOS). Los Angeles, CA: Western Psychological Services; 1999.

Maggio V, Grañana NE, Richaudeau A, Torres S, Giannotti A, Suburo AM. Behavior problems in children with specific language impairment. Journal of Child Neurology. 2014; 29(2):194–202. [PubMed: 24272522]

March JS, Parker JDA, Sullivan K, Stallings P, Conners CK. The Multidimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. Journal of the American Academy of Child & Adolescent Psychiatry. 1997; 36(4):554–565. [PubMed: 9100431]

Muratori F, Narzisi A, Tancredi R, Cosenza A, Calugi S, Saviozzi I, … Calderoni S. The CBCL 1.5–5 and the identification of preschoolers with autism in Italy. Epidemiology and Psychiatric Sciences. 2011; 20(04):329–338. [PubMed: 22201210]

Mullen, EM. Mullen scales of early learning. Circle Pines, MN: American Guidance Service; 1995.

Myers CL, Gross AD, McReynolds BM. Broadband behavior rating scales as screeners for autism? Journal of Autism and Developmental Disorders. 2014; 44(6):1403–1413. [PubMed: 24287879]

Narzisi A, Calderoni S, Maestro S, Calugi S, Mottes E, Muratori F. Child Behavior Check List 1½–5 as a tool to identify toddlers with autism spectrum disorders: A case-control study. Research in Developmental Disabilities. 2013; 34(4):1179–1189. [PubMed: 23376628]

Newcombe RG. Interval estimation for the difference between independent proportions: Comparison of eleven methods. Statistics in Medicine. 1998; 17(8):873–890. [PubMed: 9595617]

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]

Ooi YP, Rescorla L, Ang RP, Woo B, Fung DS. Identification of autism spectrum disorders using the Child Behavior Checklist in Singapore. Journal of Autism and Developmental Disorders. 2011; 41(9):1147–1156. [PubMed: 20405192]

Ooi YP, Rescorla L, Sung M, Fung DS, Woo B, Ang RP. Comparisons between autism spectrum disorders and anxiety disorders: Findings from a clinic sample in Singapore. Asia-Pacific Psychiatry. 2014; 6(1):46–53. [PubMed: 23857779]

Rescorla L, Kim YA, Oh KJ. Screening for ASD with the Korean CBCL/1½–5. Journal of Autism and Developmental Disorders. 2014 Advance online publication.

Rutter, M.; Bailey, A.; Lord, C. Social Communication Questionnaire. Los Angeles, CA: Western Psychological Services; 2003.

Rutter, M.; LeCouteur, A.; Lord, C. Autism Diagnostic Interview–Revised. Los Angeles, CA: Western Psychological Services; 2003.

Sikora DM, Hall TA, Hartley SL, Gerrard-Morris AE, Cagle S. Does parent report of behavior differ across ADOS-G classifications: Analysis of scores from the CBCL and GARS. Journal of Autism and Developmental Disorders. 2008; 38(3):440–448. [PubMed: 17619131]

So P, Greaves-Lord K, Van der Ende J, Verhulst FC, Rescorla L, de Nijs PF. Using the Child Behavior Checklist and the Teacher's Report Form for identification of children with autism spectrum disorders. Autism. 2013; 17(5):595–607. [PubMed: 22914776]

Sparrow, S.; Cicchetti, D.; Balla, D. Vineland adaptive behavior scales. 2. Circle Pines, MN: American Guidance Service; 2005.

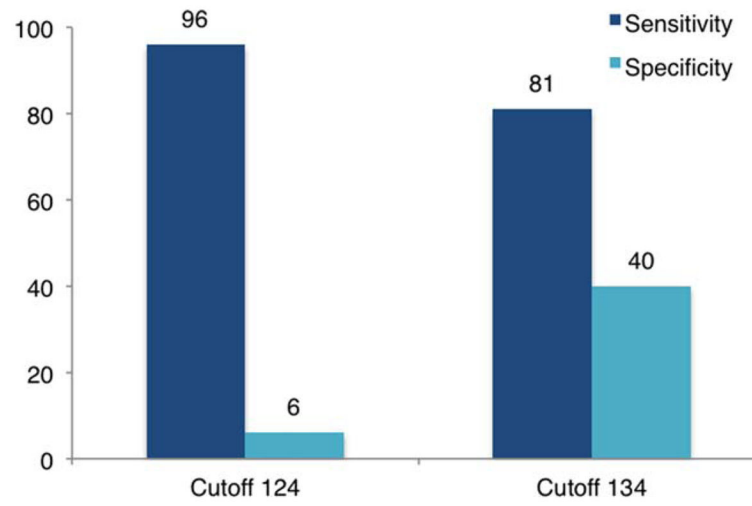Spence SH. A measure of anxiety symptoms among children. Behaviour Research and Therapy. 1998; 36(5):545–566. [PubMed: 9648330]

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

Figure 1.
Sensitivity and specificity (%) of the WTP scale in children with high EBP ($n = 80$) at cutoff 124 and 134. Abbreviation: EBP = emotional/behavioral problems.

Table 1

Sample Characteristics

Characteristic	CBCL/1.5-5			CBCL/6-18		
	ASD (*n* = 104)	Non-ASD (*n* = 57)	*t*/X²	ASD (*n* = 122)	Non-ASD (*n* = 106)	*t*/X²
Age, years, *m* (SD)	4.2 (1.1)	4.4 (0.9)	−1.4	9.4 (1.8)	9.2 (2.0)	0.8
Gender, male, *n* (%)	85 (81.7)	43 (75.4)	0.9	91 (74.6)	69 (65.1)	2.4
Nonverbal IQ, *m* (SD)	76.9 (25.7)	97.6 (20.5)	−5.6 ***	87.0 (28.3)	90.7 (19.7)	−1.2
Verbal IQ, *m* (SD)	69.5 (32.7)	92.3 (22.1)	−5.2 ***	81.2 (29.6)	91.3 (22.1)	−3.0 **
ADOS module, *n* (%)			14.5 **			9.9 **
1: Single words or less	54 (51.9)	12 (21.1)		16 (13.1)	2 (1.9)	
2: Phrase speech	27 (26.0)	25 (43.9)		9 (7.4)	8 (7.5)	
3: Fluent speech	23 (22.1)	20 (35.1)		97 (79.5)	96 (90.6)	
ADOS comparison score, *m* (SD)	7.3 (1.8)	2.3 (2.2)	12.9 ***	7.2 (2.2)	2.4 (1.8)	18.3 ***
High EBP-level, *n* (%)	34 (32.7)	12 (21.1)	2.4	45 (36.9)	35 (33.0)	0.4
Non-ASD diagnoses, *n* (%)[a]						34.3 ***
ADHD		14 (24.6)			48 (45.3)	
Intellectual disability		5 (8.8)			21 (19.8)	
Language disorder		33 (57.9)			15 (14.2)	
Emotional disorder		5 (8.8)			22 (20.8)	

Note. CBCL = child behavior checklist, ASD = autism spectrum disorder, ADOS = autism diagnostic observation schedule, EBP = emotional/behavioral problems, ADHD = attention deficit/hyperactivity disorder.

1 preschool ASD case had missing on IQ.

[a] Comparison between preschool and school-age non-ASD groups.

*
    P<0.05;

**
    P<0.01;

***
    P<0.001.

Table 2

Mean (SD) Raw Scores on the CBCL/1.5–5 (N = 161) and the CBCL/6–18 (N = 228) for Children With ASD and Non-ASD Disorders

CBCL/1.5–5	ASD (n = 104)	Non-ASD (n = 57)	F	df	η^2_P
Broadband scales *			2.79	3,153	0.05
Total problems *	63.7 (32.7)	50.4 (29.6)	6.36	1,153	0.04
Internalizing **	18.1 (11.2)	13.7 (10.2)	8.28	1,153	0.05
Externalizing	22.6 (11.3)	18.6 (11.4)	3.77	1,153	0.02
Syndrome scales *			2.25	7,149	0.10
Emotionally reactive **	5.3 (4.4)	4.1 (4.1)	7.93	1,149	0.05
Anxious/depressed	3.6 (3.1)	3.3 (3.0)	1.54	1,149	0.01
Somatic complaints	3.4 (2.9)	2.6 (2.6)	3.32	1,149	0.02
Withdrawn **	5.9 (3.3)	3.7 (3.1)	10.35	1,149	0.06
Sleep problems	4.7 (3.7)	4.1 (3.1)	0.86	1,149	0.01
Attention problems	5.2 (2.4)	4.3 (2.9)	1.04	1,149	0.01
Aggressive behavior *	17.5 (9.7)	14.4 (9.5)	4.03	1,149	0.03
DSM-oriented scales *			2.75	5,151	0.08
Affective problems	4.6 (3.1)	3.5 (3.1)	2.60	1,151	0.02
Anxiety problems *	5.1 (4.5)	4.3 (3.6)	4.30	1,151	0.03
PDP **	10.3 (5.4)	7.3 (5.0)	11.89	1,151	0.07
ADHD problems	7.5 (3.0)	6.2 (3.2)	2.78	1,151	0.02
ODD problems	5.9 (3.4)	5.0 (3.6)	3.07	1,151	0.02

CBCL/6–18	ASD (n = 122)	Non-ASD (n = 106)	F	df	η^2_P
Broadband scales ***			8.95	3,221	0.11
Total problems	62.4 (28.0)	55.2 (29.9)	3.58	1,221	0.02
Internalizing	14.4 (8.2)	12.8 (8.7)	2.54	1,221	0.01
Externalizing	14.6 (10.4)	15.0 (11.4)	0.03	1,221	0.00
Syndrome scales ***			7.75	8,216	0.22

CBCL/1.5–5	ASD (*n* = 104)	Non-ASD (*n* = 57)	*F*	df	η_P^2
Anxious/depressed	7.0 (5.3)	7.1 (5.2)	0.00	1,216	0.00
Somatic complaints	2.9 (2.9)	2.7 (2.9)	0.78	1,216	0.00
Withdrawn/depressed ***	4.5 (2.8)	3.0 (2.7)	14.98	1,216	0.06
Social problems *	7.8 (4.1)	6.6 (4.1)	5.53	1,216	0.02
Thought problems ***	7.9 (4.9)	5.2 (4.6)	17.16	1,216	0.07
Attention problems	10.9 (4.4)	9.9 (4.7)	2.09	1,216	0.01
Aggressive behavior	11.3 (8.0)	11.0 (8.1)	0.17	1,216	0.00
Rulebreaking behavior	3.3 (3.0)	4.0 (3.8)	2.34	1,216	0.01
DSM-oriented scales			0.48	6,218	0.01
Affective problems	5.2 (3.4)	4.6 (3.6)			
Anxiety problems	4.5 (3.3)	4.3 (3.2)			
Somatic problems	1.8 (2.1)	1.7 (2.2)			
ADHD problems	8.0 (3.6)	8.1 (3.9)			
ODD problems	4.4 (2.9)	4.6 (3.0)			
Conduct problems	4.9 (5.0)	5.0 (5.4)			

Note. CBCL = child behavior checklist, ASD = autism spectrum disorder, PDP = pervasive developmental problems, ADHD = attention deficit/hyperactivity, ODD = oppositional/defiant, η_P^2 = partial eta squared.

1 case excluded from MANCOVA due to missing on IQ.

*
P <0.05;

**
P <0.01;

***
P <0.001.

Table 3

Mean (SD) T Scores and Area Under the Curve (AUC) Scores for ASD Screening Scales

Preschool CBCL/1.5–5	ASD (*n* = 104)	Non-ASD (*n* = 57)	AUC	95% CI	SE
Withdrawn	68.9 (10.9)	62.0 (10.3)	0.69 ***	0.61–0.78	0.04
PDP	71.0 (11.3)	64.0 (11.3)	0.68 ***	0.59–0.76	0.05
School-age CBCL/6–18	ASD (*n* = 122)	Non-ASD (*n* = 106)	AUC	95% CI	SE
Withdrawn/depressed	64.7 (8.5)	59.8 (8.4)	0.67 ***	0.60–0.74	0.04
Thought problems	68.7 (9.5)	62.9 (10.0)	0.67 ***	0.60–0.74	0.04
Social problems	66.6 (9.4)	63.9 (9.4)	0.59 *	0.51–0.66	0.04
WTP scale	133.5 (13.9)	122.7 (15.8)	0.70 ***	0.63–0.77	0.04

Note. ASD = autism spectrum disorder, CBCL = Child behavior checklist, PDP = pervasive developmental problems, WTP = Withdrawn-Thought Problems (aggregated *T* scores), CI = confidence interval, SE = standard error.

*
P<0.05;

**
P<0.01;

***
P<0.001.

Table 4

Area Under the Curve (AUC) Scores, Sensitivity, Specificity and Positive Likelihood Ratio (95% Confidence Intervals) of the Withdrawn and WTP Scales in the Total Sample and in Subgroups

CBCL/1.5–5 Withdrawn (*T* score 62)	AUC	Sensitivity, %	Specificity,%	Likelihood ratio+
Stratification (*n* ASD/*n* Non-ASD)				
Total sample (104/57)	0.69 (0.61–0.78)	74 (64–82)	53 (39–66)	1.6 (1.2–2.1)
High EBP-level (34/12)	0.62 (0.43–0.81)	88 (72–96)	8 (0–40)	1.0 (0.8–1.2)
Low EBP-level (70/45)	0.70 (0.61–0.80)	67 (55–78)	64 (49–78)	1.9 (1.2–2.9)
Previous ASD diagnosis (52)	0.74 (0.65–0.84)	81 (67–90)		1.7 (1.3–2.3)
No previous ASD diagnosis (52)	0.64 (0.54–0.75)	67 (53–79)		1.4 (1.0–2.0)

CBCL/6–18 WTP (Aggregated *T* score 124)	AUC	Sensitivity,%	Specificity,%	Likelihood ratio+
Stratification (*n* ASD/*n* Non-ASD)				
Total sample (122/106)	0.70 (0.63–0.77)	78 (69–85)	55 (45–64)	1.7 (1.4–2.2)
High EBP-level (45/35)	0.62 (0.49–0.74)	96 (84–99)	6 (1–13)	1.0 (0.9–1.1)
Low EBP-level (77/71)	0.79 (0.72–0.86)	68 (56–78)	79 (67–87)	3.2 (2.0–5.1)
No ID (93/85)	0.73 (0.66–0.81)	80 (70–87)	56 (45–67)	1.8 (1.4–2.4)
ID (29/21)	0.59 (0.42–0.76)	72 (53–87)	48 (26–70)	1.4 (0.9–2.2)
Previous ASD diagnosis (70)	0.70 (0.62–0.77)	80 (68–88)		1.8 (1.4–2.2)
No previous ASD diagnosis (52)	0.71 (0.63–0.79)	75 (61–86)		1.7 (1.3–2.2)

Note. CBCL = Child behavior checklist, ASD = autism spectrum disorder, WTP = Withdrawn-Thought Problems, EBP = emotional/behavioral problems, ID = intellectual disability.