Receiver operating characteristic curve analysis of the Child Behavior Checklist and Teacher’s Report Form for assessing autism spectrum disorder in preschool-aged children

Tiago S Bara,1,2 Antonio C Farias,1,3,4 Erico PG Felden2 Mara L Cordeiro1,2,6

1Neurosciences Core, Pelé Pequeno Príncipe Research Institute, 2Facultades Pequeno Príncipe, 3Department of Neuropediatrics, Children’s Hospital, Pequeno Príncipe, 4School of Medicine, University Positivo, Curitiba, 5Centro de Ciências da Saúde e do Esporte (CEFID) da Universidade do Estado de Santa Catarina, Florianópolis, Brazil; 6Department of Psychiatry and Biobehavior Sciences, David Geffen School of Medicine, Semel Institute for Neurosciences and Human Behavior, University of California Los Angeles, Los Angeles, CA, USA

Background: Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by social, behavioral, and communication impairments with an estimated prevalence of 1 in 68 school-aged children. There is a need for objective and easily applicable instruments for early identification of autistic children to enable initiation of early interventions during a very sensitive period of brain development and, consequently, optimize prognosis. Here, we tested the utility of the Child Behavior Checklist (CBCL) and the Caregiver-Teacher’s Report Form (C-TRF) scales for assessing ASD in Brazil, where ASD screening research is emergent.

Subjects and methods: A total of 70 children (2–5 years old, both sexes) were enrolled, including an ASD group (n = 39) and a non-ASD control group (n = 31). The preschool versions of the CBCL and C-TRF were applied. The CBCL and C-TRF results were compared between the ASD and non-ASD control groups with Mann–Whitney U tests and receiver operating characteristic analyses.

Results: The CBCL and C-TRF were found to have moderate accuracy for the dimensions withdrawn and autism spectrum problems, and to correlate with each other.

Conclusion: The CBCL and C-TRF may aid in early ASD detection.

Keywords: screening, instruments, AUC, early intervention

Introduction

Autism spectrum disorder (ASD) encompasses a range of neurodevelopmental phenotypes characterized by social, behavioral, and communication impairments.1,2 ASD diagnosis remains reliant on a clinical assessment of behavioral traits without corroborative biomarkers.3,4 The diagnostic criteria for ASD established by the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5) encompass a dyad of symptoms that should be present from early childhood and limit or interfere with daily functioning: 1) a persistent deficit in social communication and social interactions and 2) restricted and repetitive patterns of behavior, interests, or activities.5,6

A growing body of research has addressed the importance of early detection and screening of ASD in toddlers to improve outcomes.7–10 The American Academy of Neurology recommends that ASD diagnoses be made on the basis of two levels of investigation, wherein level 1 identifies any type of atypical development as a risk factor for ASD and level 2 involves more in-depth investigation and differentiation of ASD from other developmental disorders.11 Level 1 instruments include the Modified...
Checklist for Autism in Toddlers,\textsuperscript{15} the Children’s Social Behavior Questionnaire,\textsuperscript{16} the Child Behavior Checklist (CBCL), and the Teacher’s Report Form (TRF).\textsuperscript{17} Level 2 instruments include the Autism Diagnostic Observation Schedule-Generic, the Gilliam Autism Rating Scale,\textsuperscript{18} the Social Communication Questionnaire,\textsuperscript{19} and the Social Responsiveness Scale.\textsuperscript{20}

The CBCL and TRF are widely used in research for evaluation of emotional and behavioral problems; they are appreciated for their high sensitivity and specificity relative to other level 1 instruments.\textsuperscript{17,21} They consist of questionnaires to be completed by a child, his or her parent/guardian, and his or her teacher/caregiver. Both scales evaluate social skills and behavioral problems in children 1.5–18 years of age.\textsuperscript{17,21} They identify internalizing and externalizing behaviors as well as possible indicators of mental disorders.\textsuperscript{22,23} A small number of studies have examined the potential of these instruments as ASD screening tools.\textsuperscript{10,12,21}

Only limited studies in Brazil have probed the potential utility of using the CBCL to assess ASD symptoms,\textsuperscript{24,25} and there has not yet been any systematic studies in Brazil to test the applicability of the CBCL and TRF for ASD screening, particularly among preschool-aged children. Moreover, the substantially lower ASD prevalence estimates reported for school-aged children in Brazil (0.3%, 1/368),\textsuperscript{26} relative to that in the USA (1.5%, 1/68),\textsuperscript{27} suggests that ASD may be substantially underestimated in Brazil. In this context, there is a need for objective, reliable, and easy-to-interpret instruments that can help achieve broader early diagnosis of ASD among young children in Brazil.\textsuperscript{28,29} Furthermore, because ASD diagnosis is made primarily by clinical observation, without any definitive or supporting laboratory or imaging tests, it is important to gather information from multiple informants, such as parents and caregiver/teachers, and it is vital that the instruments used record observers’ opinions accurately.

The aim of this study was to test the hypothesis that the CBCL and Caregiver-Teacher Report Form (C-TRF) are reliable tools for distinguishing ASD preschoolers from a heterogeneous sample of similarly aged children without ASD. The results of the two groups were compared statistically, and the diagnostic accuracy of the two instruments was assessed with area under the curve (AUC) analysis of receiver operating characteristic (ROC) curves.

**Subjects and methods**

**Participants**

Seventy children (9 girls and 61 boys, 2–5 years old) participated in this research, including an ASD group (n=39) and a non-ASD control group (n=31). The ASD group (5 girls and 34 boys) was composed entirely of children diagnosed with ASD according to the DMS-5 criteria.\textsuperscript{1} The control group (4 girls and 27 boys) was composed of children with social communication disorder (SCD, n=11), language developmental disorder (LDD, n=13), attention deficit/hyperactivity disorder (ADHD, n=3), anxiety disorder (n=1), and obsessive–compulsive disorder (n=1), as well as typically developing children (n=2). The instruments were applied in this study before the children had an official diagnosis when each child first entered our center. The aforementioned official diagnoses were obtained subsequently through a full neuropsychiatric team evaluation. Ethical approval for the study was provided by the Ethics Committee on Human Research at the Pequeno Principe Children’s Hospital in Curitiba, Brazil, for all aspects of the research protocol (registration number: 67276717.7.0000.0097). Written informed consent for study participation was obtained from a parent or legal guardian of each child who participated in this study.

**Instruments**

**CBCL**

Parents (usually the mother) completed the standardized Brazilian version of the CBCL 1.5–5 years, previously shown to have semantic principles consistent with the original CBCL, as well as excellent validity and reliability.\textsuperscript{30} The instrument includes a social competence and adaptation domain and a behavioral problems domain.\textsuperscript{17}

**C-TRF**

The standardized Brazilian version of the C-TRF was completed by a teacher or caregiver of each participating child. The reliability and validity of the C-TRF were established previously.\textsuperscript{31}

**Evaluation procedure**

The children and their parents completed the evaluations. Teacher/caregiver questionnaires were distributed by teacher/caregiver referral managers and they were returned to our research center on the day of each child’s on-site evaluation. Raw CBCL and C-TRF scores were converted to age- and gender-standardized T-scores in Achenbach System of Empirically Based Assessment software,\textsuperscript{11} such that the average score for each age and gender was associated with a T-score of 50 and an SD of 10.

**Data analysis**

The data were analyzed in SPSS for Windows\textsuperscript{®} software version 21.0. Descriptive analyses for central tendency and dispersion of all analyzed variables were calculated.
Data were compared across the two groups of children with Mann-Whitney U tests. To obtain summative statistical measures of the tests’ diagnostic discrimination abilities, the scales were subjected to AUC analysis of ROC curves, the gold standard of ASD diagnosis. We calculated 95% CIs for ASD diagnosis from the ROC analysis results based on cutoff points established by binary logistic regression with a significance level of 5%. The ROC analysis results were interpreted as follows: AUC <0.70, low diagnostic accuracy; AUC in the range of 0.70–0.90, moderate diagnostic accuracy; and AUC ≥0.90, high diagnostic accuracy. Correspondence between the scales was assessed with a Spearman correlation test. For correlational test result interpretation, rho (\( \rho \)) values were interpreted as follows: \( \rho <0.20 \), very weak; \( 0.20–0.39 \), weak; \( 0.40–0.59 \), moderate; \( 0.60–0.79 \), strong; and \( \geq 0.80 \), very strong.

**Results**

**Demographic characteristics**
The mean ages of the ASD (3.41±0.94 years) and control (3.68±1.14 years) groups were similar (\( p=0.123 \)), as were the groups’ gender ratios (12.8% and 12.9% girls, respectively; \( p=0.990 \)). The girls and boys were similar in terms of age (\( p=0.791 \)), scores for all dimensions of the CBCL (\( p>0.05 \)), except the ADHD dimension (boys mean score, 59.21±7.11; girls mean score, 53.33±4.24; \( p=0.021 \)) and the mean C-TRF scores (\( p>0.05 \)).

**CBCL**
As illustrated in Figure 1, the ASD group had higher mean scores than the control group on five CBCL dimensions: withdrawn (W; \( p<0.001 \)), depressive problems (DP; \( p=0.016 \)), autism spectrum problems (ASP; \( p<0.001 \)), internalizing problems (IP; \( p=0.007 \)), and total problems (TP; \( p=0.011 \)). As reported in Table 1, the cutoff points based on ROC analyses for these dimensions showed significant predictive power of the CBCL for ASD, with the ASP and W dimensions exhibiting high accuracy (AUC graphs in Figure 2) and the DP, IP, and TP dimensions exhibiting low accuracy. The AUCs, \( p \)-values, cutoff points optimized for sensitivity and specificity, 95% CIs, sensitivities, and specificities obtained for the five significant CBCL dimensions are reported in Table 1.

![Figure 1](image-url)  
**Figure 1** Group differences in CBCL (1.5–5 years) dimensions scores.  
**Notes:** Relative to the non-ASD comparison control group, the ASD group had significantly higher scores in the ASP and W dimensions as well as in the DP, IP, and TP dimensions. \( *p<0.05. \)  
**Abbreviations:** ASD, autism spectrum disorder; ASP, autism spectrum problems; CBCL, Child Behavior Checklist; DP, depressive problems; IP, internalizing problems; TP, total problems; W, withdrawn.
As shown in Figure 3, the ASD group obtained higher mean scores than the control group on the emotionally reactive (ER; \( p = 0.013 \)), W (\( p = 0.001 \)), attention problems (AP; \( p = 0.028 \)), ASP (\( p < 0.001 \)), IP (\( p < 0.006 \)), and TP (\( p < 0.040 \)) dimensions of the C-TRF. ASD was associated with T-scores.

Table 1

<table>
<thead>
<tr>
<th>CBCL dimension</th>
<th>AUC</th>
<th>p-value</th>
<th>Cutoff point</th>
<th>95% CI</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotionally reactive</td>
<td>–</td>
<td>0.196</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Anxious/depressed</td>
<td>–</td>
<td>0.813</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Somatic complaints</td>
<td>–</td>
<td>0.223</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Withdrawal</td>
<td>0.768</td>
<td>&lt;0.001*</td>
<td>63</td>
<td>0.65–0.86</td>
<td>71.0</td>
<td>69.2</td>
</tr>
<tr>
<td>Sleep problems</td>
<td>–</td>
<td>0.275</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Attention problems</td>
<td>–</td>
<td>0.128</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Aggressive behavior</td>
<td>–</td>
<td>0.324</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Depressive problems</td>
<td>0.666</td>
<td>0.009*</td>
<td>56</td>
<td>0.54–0.77</td>
<td>67.7</td>
<td>61.5</td>
</tr>
<tr>
<td>Anxiety problems</td>
<td>–</td>
<td>0.171</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Autism spectrum problems</td>
<td>0.781</td>
<td>&lt;0.001*</td>
<td>70</td>
<td>0.67–0.87</td>
<td>80.6</td>
<td>71.8</td>
</tr>
<tr>
<td>Attention deficit/hyperactivity problems</td>
<td>–</td>
<td>0.236</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Oppositional defiant problems</td>
<td>–</td>
<td>0.976</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Internalizing problems</td>
<td>0.687</td>
<td>0.002*</td>
<td>61</td>
<td>0.56–0.79</td>
<td>71.0</td>
<td>71.8</td>
</tr>
<tr>
<td>Total problems</td>
<td>0.678</td>
<td>0.005*</td>
<td>56</td>
<td>0.55–0.78</td>
<td>51.6</td>
<td>87.2</td>
</tr>
</tbody>
</table>

Note: \(^* p<0.05\).

Abbreviations: AUC, area under the curve; CBCL, Child Behavior Checklist; ROC, receiver operating characteristic.

Correlation between CBCL and C-TRF results

A positive correlation was observed between the CBCL and C-TRF for most of the dimensions, with the exceptions of DP (\( p = 0.632 \)) and attention deficit/hyperactivity problems (ADHP) (\( p = 0.177 \)). The strengths of the correlations for the ASP (\( p < 0.001, r = 0.481 \)), IP (\( p < 0.001, r = 0.418 \)), EP (\( p < 0.001, r = 0.440 \)), and TP (\( p < 0.001, r = 0.401 \)) dimensions were moderate according to Evans classification rubric.\(^{31}\)

Discussion

Few instruments have been adapted and validated for ASD detection in Brazilian children. This study was the first in Brazil to investigate the discriminative capacity of the CBCL (1.5–5 years) and C-TRF for identification of ASD signs in

C-TRF

As shown in Figure 3, the ASD group obtained higher mean scores than the control group on the emotionally reactive (ER; \( p = 0.013 \)), W (\( p < 0.001 \)), attention problems (AP; \( p = 0.028 \)), ASP (\( p < 0.001 \)), IP (\( p < 0.006 \)), and TP (\( p < 0.040 \)) dimensions of the C-TRF. ASD was associated with T-scores >70 on the ASP and W dimensions. As reported in Table 2, the cutoff points based on ROC analysis for these six C-TRF dimensions showed significant predictive power for ASD, with the ASP and W dimensions exhibiting high accuracy (AUC graphs in Figure 4) and the others exhibiting low accuracy. The AUCs, \( p \)-values, cutoff points optimized for sensitivity and specificity, 95% CIs, sensitivities, and specificities obtained for the six significant C-TRF dimensions are reported in Table 2.
preschool-aged children. The results we obtained are consistent with prior findings in indicating that the W and ASP dimensions of these instruments are promising discriminatory factors for children with ASD. Highlights of our research include the use of inventories for both parents and teachers and the use of T-score cutoffs that are appropriate according to the developers of the instruments.17

Good instrument sensitivity is conducive to screening applications, for which false positives are more acceptable than misses, as opposed to diagnostically determinant

Table 2 ROC analysis for screening ability of the C-TRF

<table>
<thead>
<tr>
<th>C-TRF dimension</th>
<th>AUC</th>
<th>p-value</th>
<th>Cutoff point</th>
<th>95% CI</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotionally reactive</td>
<td>0.675</td>
<td>0.007</td>
<td>63</td>
<td>0.55–0.78</td>
<td>70.0</td>
<td>60.5</td>
</tr>
<tr>
<td>Anxious/depressed</td>
<td>–</td>
<td>0.214</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Somatic complaints</td>
<td>–</td>
<td>0.183</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Withdrawn</td>
<td>0.763</td>
<td>&lt;0.001*</td>
<td>66</td>
<td>0.64–0.86</td>
<td>80.6</td>
<td>60.5</td>
</tr>
<tr>
<td>Attention problems</td>
<td>0.654</td>
<td>0.018*</td>
<td>64</td>
<td>0.53–0.76</td>
<td>71.0</td>
<td>63.2</td>
</tr>
<tr>
<td>Aggressive behavior</td>
<td>–</td>
<td>0.499</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Depressive problems</td>
<td>–</td>
<td>0.229</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Anxiety problems</td>
<td>–</td>
<td>0.084</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Autism spectrum problems</td>
<td>0.790</td>
<td>&lt;0.001*</td>
<td>62</td>
<td>0.67–0.88</td>
<td>53.3</td>
<td>94.7</td>
</tr>
<tr>
<td>Attention deficit/hyperactivity problems</td>
<td>–</td>
<td>0.136</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Oppositional defiant problems</td>
<td>–</td>
<td>0.363</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Internalizing problems</td>
<td>0.694</td>
<td>0.002*</td>
<td>68</td>
<td>0.57–0.80</td>
<td>83.9</td>
<td>50.0</td>
</tr>
<tr>
<td>Total problems</td>
<td>0.644</td>
<td>0.029*</td>
<td>64</td>
<td>0.51–0.75</td>
<td>54.8</td>
<td>73.7</td>
</tr>
</tbody>
</table>

Note: *p<0.05.

Abbreviations: AUC, area under the curve; C-TRF, Caregiver-Teacher Report Form; ROC, receiver operating characteristic.
applications.11 Our ROC analysis indicated that the ASP dimension of the CBCL showed moderate accuracy (AUC = 0.79), but a sensitivity of only 53.3%, which suggests reliance on it would produce a large number of false negatives (46.7%). Others have obtained better sensitivity values with a lower percentage of false negatives for this dimension than we found here. For example, in a study comparing an ASD group (n = 101) with a group of children with psychiatric disorders other than ASD, SCD, or LDD (n = 95), Muratori et al10 found an ideal cutoff point for the Pervasive Developmental Problems (PDP) dimension (analogous to the current ASP dimension) at a T-score of 65 with a sensitivity of 85%, specificity of 60%, and moderate accuracy (AUC = 0.813) for identifying children with ASD. In a study with a sample of 94 children (47 with ASD), Narzisi et al11 obtained a T-score cutoff point for the PDP dimension of 69, with a sensitivity of 85%, specificity of 83%, and high accuracy (AUC = 0.914) for identifying children with ASD. Finally, Sikora et al12 obtained a sensitivity of 79.75 with a specificity of 42.0% for the identification of children with ASD with the PDP dimension.

The inclusion of a substantial number of children with SCD in our comparison control group (11/31) could have expanded the non-ASD score characteristics beyond neurotypical range into ASD range. Indeed, Havdahl et al9 suggested that the CBCL may be inadequate for distinguishing ASD from other language or communication deficits, especially SCD. The SCD diagnosis in the DSM-5 applies to individuals who have problems with the social use of language without the restricted/repetitive behaviors characteristic of ASD. The CBCL may miss ASD cases due to a lack of items that assess the presence of restricted/repetitive behavioral patterns, the behavioral criterion that distinguishes ASD from SCD.

Relative to the sensitivity of the CBCL in this study, the C-TRF yielded a better sensitivity (80.6%) and specificity (71.8%) balance, together with moderate accuracy and, thus, a lower false-negative rate. These results put the ASP dimension T-score cutoff in the borderline range,17 indicating that the C-TRF may be more useful than the CBCL for identifying children with likely ASD and for distinguishing ASD from other disorders.

Despite observing a positive correlation overall between the two instruments, we observed only a moderate correlation for the ASP dimension and a weak correlation for the W dimension. Hence, it may be more effective to consider the outcomes of the ASP dimension for both instruments (CBCL and C-TRF) when screening ASD. So et al have made a strong case for the greater screening power that can be achieved by combining scales.21 Using versions of the CBCL and TRF for 6–18-year-old subjects in a large study (458 ASD and 1,109 non-ASD participants), they found that joint application of the instruments increased positive and negative predictive values substantially, relative to either of them alone.

This study has some limitations that should be considered. First, because we examined a referred, rather than a random, sample, the results might not be generalized to the general population. Second, our comparison control group was composed mostly of children with a variety of other behavioral disorders characterized by traits that overlap with ASD, including SCD,
ADHD, anxiety disorder, obsessive–compulsive disorder, and LDD. Third, we obtained high false-negative rates (i.e., 100–specificity), which may be due to a lack of ASD focal points in the inventories or feature overlap between ASD and other social/communication disorders without adequate assessment of ASD-associated restricted/repetitive behaviors. Notably, the diagnostic criteria for SCD versus ASD have not yet been fully explored with Achenbach System of Empirically Based Assessment instruments.

Despite the aforementioned limitations, our findings support the notion that use of the CBCL and C-TRF, especially together, may aid in screening for ASD among preschoolers. This is the first study to evaluate the screening potential of the CBCL (1.5–5 years) and C-TRF under the scope of DSM-5 in a well-characterized preschool population of children diagnosed with ASD. The results suggest that the ASP dimension of these scales may distinguish children with ASD from children with other disorders. Notwithstanding, the results also suggest that the CBCL and C-TRF continue to follow the configuration of the DSM-IV-TR diagnostic criteria,31 despite updates to the DSM-5 nomenclature. Their validity, particularly for newly designated diagnoses such as SCD, may be improved by the revision or addition of items that bring the contents into closer alignment with the DSM-5. Future studies may evaluate the predictive capacity of these instruments for the new SCD diagnosis, which is important to prevent ASD misdiagnosis in children with SCD.

**Conclusion**

The findings of this study support the hypothesis that the C-TRF and CBCL (1.5–5 years) may be useful for ASD screening in preschool-aged children. In particular, our ROC analysis showed that the ASP and W dimensions of these instruments have moderate accuracy for detection of ASD. Screening of preschoolers for ASD with these instruments may reduce undiagnosis and, thus, lead to the delivery of early intervention to more children with ASD, resulting in better outcomes.

**Acknowledgment**

This study was funded by the SETI State Secretary of Science, Technology and Superior Education of Paraná (SETI, grant# conv. 07/14).

**Author contributions**

MLC and TSB conceptualized and designed the study protocol. ACF and TSB did data acquisition. EPGF did all the statistical analyses. MLC and TSB interpreted the findings and wrote the manuscript. MLC performed a critical revision of the manuscript for important intellectual content and was responsible for submitting the final approved manuscript. All authors helped with revising the paper and agreed to be accounted for all aspects of the work.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**