



Archived by Flinders University

This is the peer reviewed version of the following article:

Lucas, C. A., Brewer, N., & Young, R. L. (2022). Pitfalls When Using Area Under the Curve to Evaluate Item Content for Early Screening Tests for Autism. In *Journal of Psychoeducational Assessment*. SAGE Publications.

which has been published in final form at

<https://doi.org/10.1177/07342829211067128>

Any re-use terms are restricted to non-commercial and no derivative uses. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Copyright The Author(s) 2022

Pitfalls When Using AUC to Evaluate Item Content for Early Screening Tests for Autism

Carmen A. Lucas, Neil Brewer, and Robyn L. Young

Flinders University

Author note:

Carmen A. Lucas, Neil Brewer, and Robyn L. Young, College of Education, Psychology and Social Work, Flinders University. Correspondence concerning this article should be addressed to Carmen Lucas, College of Education, Psychology and Social Work, Flinders University, GPO Box 2100 Adelaide, South Australia 5001. Email:

carmen.lucas@flinders.edu.au

Acknowledgements: We are very grateful to Yong-Hwee Nah and Rose Nevill for providing us with their data files. Nah et al.'s (2018) dataset is available at <https://osf.io/utxrc/>

Declaration of Conflicts of Interest: The university with which the authors are affiliated receives royalties from the sale of the ADEC.

Funding: This work was supported by the Australian Research Council [DP 190100162] and the Hamish Ramsay Fund.

Abstract

Evaluations of early screening tests for autism commonly rely on receiver operating characteristic (ROC) analysis and comparisons of area under the curve (AUC). Whether AUC differs significantly from chance or between test items is not always assessed. Two recent and independent evaluations of the Brief Autism Detection in Early Childhood (BADEC) constructed a short form by selecting the five items with the highest AUC values, leading to inconsistencies regarding appropriate item content (Nah et al., 2018; Nevill et al., 2019). Using significance testing to compare AUC values for each test item from each dataset, we demonstrate which items justify inclusion in the BADEC, which items can be ruled out, and highlight key factors influencing AUC significance testing outcomes.

Keywords: Autism screening; ROC analysis; AUC comparisons; ADEC; BADEC

Pitfalls When Using AUC to Evaluate Item Content for Early Screening Tests for Autism

It is accepted that developmental outcomes for individuals with autism spectrum disorder (ASD) will benefit from diagnosis at a young age and subsequent early intervention (e.g., Schreibman, 2000). To facilitate widespread screening for ASD in children aged 12–36 months, researchers have developed Level 1 screeners for screening children in the general population who exhibit signs of autism (e.g., Dietz et al., 2006) and Level 2 screeners to distinguish ASD from other developmental disabilities (e.g., Stone et al., 2004; Young, 2007). Here we address an important issue regarding the evaluation of screeners: namely, the consequences of failing to determine whether items differ meaningfully from one another in their ability to detect the condition of interest.

An important part of developing a screener is to collect data evaluating its diagnostic accuracy; through this process the screener's ability to differentiate between those with and without the target condition is established. It is common to evaluate the discriminative performance of early screening test items, or whole tests, using receiver operating characteristics (ROC) analysis. ROC analysis involves plotting the proportion of individuals among those with the relevant condition who screen positive on the item or test (i.e., sensitivity or S_E) against the proportion of individuals without the condition who screen negative (i.e., specificity or S_P) across the range of possible cutoff scores. The result is a ROC curve which plots S_E values against $1-S_P$ values (i.e., hits against false alarms). An area under the curve (AUC) statistic is then used to assess the test's (or item's) discriminative performance. AUCs of .50 and 1 indicate chance and perfect discrimination, respectively. A significance test is necessary to formally evaluate AUC (e.g., whether performance differs meaningfully from chance or another AUC value); however, this approach is not commonly used in practice.

Concerns regarding the inconsistent use of significance testing to evaluate AUC in the context of autism screener development were raised by Brewer et al.'s (2020) review of Level 2 ASD screeners, using as an example two evaluations of brief versions of the Autism Detection in Early Childhood (ADEC; Young, 2007). Brief screeners—comprised of key items, often from a more comprehensive assessment tool—allow quicker assessments that are able to be more widely implemented. Nah et al. (2018) and Nevill et al. (2019) constructed short forms of the ADEC (Brief Autism Detection in Early Childhood; BADEC) by administering the full ADEC to infants with or without ASD and selecting the five items with the highest AUC. Three items—response to name, reciprocity of smile and following verbal commands—were included in both BADEC versions, while two items differed across versions. A lack of meaningful AUC comparisons between items prevents firm conclusions regarding the most discriminating items. Consequently, we used significance testing to re-evaluate the evidence for BADEC item selection in the Nah et al. and Nevill et al.¹ datasets and contrasted our results with the original conclusions reached by selecting the five items with the highest AUC.

Method

Participants

Nah et al.'s (2018) sample included 270 children aged 12–36 months ($M = 25.4$, $SD = 7.0$). Based on a best estimate clinical (BEC) DSM-5 diagnosis, 106 had ASD, 86 were non-typically developing (non-TD) and 78 were considered typically developing (TD).

Characteristics of the non-TD children included language delay, hearing impairment and learning difficulty. As Nah et al.'s focus was the ability of the BADEC to discriminate ASD from other developmental disorders, we only used the ASD and non-TD children's data in our re-analyses. Nevill et al.'s (2019) sample comprised 110 children aged 14–36 months (M

¹ We are very grateful to Yong-Hwee Nah and Rose Nevill for providing us with their data files for reanalysis.

= 28.8, $SD = 5.4$). Based on assessments from a multidisciplinary team at a hospital pediatric center, 49 children had a diagnosis of ASD and 61 had ASD ruled out. Missing data for a small number of participants in the dataset provided to us meant our evaluations of the ADEC items are based on the slightly smaller sample of 107 children, 48 and 59 with and without ASD, respectively.

The ADEC

The ADEC is a 16-item behavioral checklist developed to detect pre-verbal behaviors predicting autism in children under the age of three (Young, 2007); it is administered in 10–15 minutes by a trained assessor who prompts 16 developmentally appropriate behaviors from the child being assessed (see Table 1). The child is scored on each behavior from 0 (appropriate) to 2 (inappropriate), allowing a maximum score of 32. Several converging validity assessments have been reported (Dix et al., 2015; Hedley et al., 2015; Nah, Young, & Brewer, 2014; Nah, Young, Brewer, & Berlinger, 2014).

Results and Discussion

Using the `roc` and `auc.ci` functions in the `pROC` package (version 1.16.2; Robin et al., 2011) in R (version 3.6.1; R Development Core Team, 2019), we recalculated the AUC values and 95% confidence intervals for each ADEC item in the Nah et al. (2018) and Nevill et al. (2019) datasets (see Table 1).² Using the `roc.test` function, AUC values were compared with paired ROC tests using 10,000 bootstrap samples; the R code for constructing and comparing the AUCs, as well as the significance values for each AUC comparison can be found in Supplemental Material (Tables S1–S3, pp. 1–3).

In Nah et al. (2018), the items with the highest AUCs comprising the BADEC were response to name, reciprocity of smile, following verbal commands, joint attention/social

² Discrepancies between the AUC values reported in the original articles—in which the method for calculating AUC was not described—and those reported here were all $< .01$ and too small to meaningfully affect results.

referencing and use of gestures; in Nevill et al. (2019), the items were again response to name, reciprocity of smile and following verbal commands, accompanied by gaze monitoring and task switching instead of joint attention/social referencing and use of gestures. In both datasets, the five items comprising the BADEC had AUC values that did not differ significantly from one another ($ps > .134$ and $> .128$ in Nah et al. and Nevill et al., respectively). In the Nah et al. dataset two additional items (functional play and gaze monitoring) had AUC values that were not significantly different from any of their BADEC items ($ps > .119$). Based on the strategy of selecting the most discriminating items, significance testing suggests that seven items could have been used interchangeably in Nah et al.'s BADEC. Moreover, since one of the additional candidates (i.e., gaze monitoring) was used in Nevill et al.'s BADEC, four rather than three items were consistent contenders for inclusion in a 5-item short-form.

A further four items (i.e., task switching, joint attention/social referencing, functional play and use of gestures) had (a) AUC values in one dataset that were not significantly different from the highest scoring item ($ps > .134$) and (b) AUC values in the other dataset that were significantly lower than the highest one or two scoring items ($ps < .037$) but none of the other BADEC items ($ps > .104$). These items could be considered as the next strongest contenders for rounding out the BADEC. Only three items in Nah et al.'s (2018) dataset (ritualistic play, anticipation of social advances and nestling) and two items in Nevill et al.'s (2019) dataset (ritualistic play and nestling) had significantly lower AUC values than those of all the test items comprising the respective BADECs evaluated ($ps < .008$ in Nah et al. and $< .038$ in Nevill et al.). Thus, several additional items not included in either BADEC version (imitation, eye contact, different types of play and reaction to common sounds) cannot be ruled out as contenders for inclusion in a 5-item short-form. Table 2 provides a summary of the strongest through to the weakest BADEC candidates.

For both datasets substantial AUC differences were necessary for the difference to be significant— $\approx .08$ in Nah et al. (2018) and $\approx .11$ in Nevill et al. (2019)—suggesting that further testing with more robust sample sizes might reduce the number of BADEC contenders. To appreciate the AUC difference that would be significant for different sample sizes, we compared AUC when doubling, tripling and quadrupling the Nah et al. and Nevill et al. datasets. In one example, increasing the N to ≈ 400 —by doubling the Nah et al. dataset and quadrupling the Nevill et al. dataset—decreased the range for a significant AUC difference to $\approx .06$ (see Tables S4 & S5 in Supplemental Material, pp. 4–5). However, one cannot simply use sample size to infer whether two AUC values are likely to differ significantly. The level of inter-item correlation affects whether a difference of the same magnitude will reach significance (Robin et al., 2011). To illustrate, we compared AUC values in mock datasets where the AUC values were the same as the original dataset but inter-item correlations were optimized (see Supplemental Material Tables S10 & S11, pp. 10 & 11; for inter-item correlations in the original and mock datasets see Supplemental Material Tables S6–S9, pp. 6–9). When inter-item correlations were higher, smaller differences in AUC were necessary for the difference to be significant.

Conclusion

The current investigation reinforces the importance of evaluating the results of ROC analysis with significance testing. Our re-analyses of Nah et al.'s (2018) and Nevill et al.'s (2019) data revealed that seven and five of the most discriminating ADEC items, respectively, did not differ meaningfully from one another and that four items were in both of these groups across studies. Compared to the strategy of simply identifying the five items with the highest AUC, which suggested that three items were consistently among the most discriminative ADEC items, the approach of comparing AUC with significance testing showed that four items fell into this category. When developing a five-item instrument, the

difference between identifying four instead of three items as consistently amongst the most discriminating across studies seems non-trivial.

The current example emphasizes the important information that can come to light when using significance testing to evaluate the most discriminating items in a diagnostic test. However, the results also have implications for future work involving other types of AUC comparison. For example, when evaluating which test items are appropriate to include in an assessment instrument, researchers may wish to establish that all of their items perform better than chance levels. The current work reinforces that it would be best to use a significance test rather than simply relying on the AUC values being higher than .50. Similarly, since the minimum acceptable AUC of a test is often considered to be .70 (e.g., Compton et al., 2006), the AUC of the combined items should be measured against this threshold with a significance test. And when comparing the discriminative performance of two different assessment instruments, a significance test is the only way to determine whether one test is performing better than the other.

Our results also reinforce that sample size needs to be considered when interpreting significance test outcomes. In addition to re-evaluating the most discriminating items in Nah et al. (2018) and Nevill et al. (2019), significance testing also facilitated delineation of which additional items could be considered contenders for inclusion in the BADEC. The large number of items that could not be definitively ruled out as candidates emphasized that, ideally, item discriminability should be compared with sufficiently robust samples for moderately sized AUC differences to be meaningful. Although it is challenging to obtain sufficiently powered samples in research involving many clinical populations, this should be an objective of future research concerning the development of ASD screeners.

References

- Brewer, N., Young, R. L., & Lucas, C. A. (2020). Autism screening in early childhood: Discriminating autism from other developmental concerns. *Frontiers in Neurology, 11*, 594381. <https://doi.org/10.3389/fneur.2020.594381>
- Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*(2), 394–409. <https://doi.org/10.1037/0022-0663.98.2.394>
- Dietz, C., Swinkels, S., van Daalen, E., van Engeland, H., & Buitelaar, J. K. (2006). 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, 36*(6), 713–22. <https://doi.org/10.1007/s10803-006-0114-1>
- Dix, L., Fallows, R., & Murphy, G. (2015). Effectiveness of the ADEC as a Level 2 screening test for young children with suspected autism spectrum disorders in a clinical setting. *Journal of Intellectual and Developmental Disability, 40*(2), 179–188. <https://doi.org/10.3109/13668250.2015.1014323>
- Hedley, D., Nevill, R. E., Monroy-Moreno, Y., Fields, N., Wilkins, J., Butter, E., & Mulick, J. A. (2015). Efficacy of the ADEC in identifying autism spectrum disorder in clinically referred toddlers in the US. *Journal of Autism and Developmental Disorders, 45*(8), 2337–2348. <https://doi.org/10.1007/s10803-015-2398-5>
- Nah, Y. H., Young, R. L., & Brewer, N. (2018). Development of a brief version of the Autism Detection in Early Childhood (BADEC). *Autism, 23*(2), 494–502. <https://doi.org/doi.org/10.1177/1362361318757563>

- Nah, Y. H., Young, R. L., & Brewer, N. (2014). Using the Autism Detection in Early Childhood (ADEC) and Childhood Autism Rating Scales (CARS) to predict long term outcomes in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 44(9), 2301–2310. <https://doi.org/10.1007/s10803-014-2102-1>
- Nah, Y. H., Young, R. L., Brewer, N., & Berlinger, G. (2014). Autism Detection in Early Childhood (ADEC): Reliability and validity data for a Level 2 screening tool for autistic disorder. *Psychological Assessment*, 26(1), 215–226. <https://doi.org/10.1037/a0034472>
- Nevill, R. E., Hedley, D., & Uljarević, M. (2019). Brief Report: Replication and validation of the Brief Autism Detection in Early Childhood (BADEC) in a clinical sample. *Journal of Autism and Developmental Disorders*, 49(11), 4674–4680. <https://doi.org/10.1007/s10803-019-04153-3>
- R Development Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12(1), 77–84. <https://doi.org/10.1186/1471-2105-12-77>
- Schreibman, L. (2000). Intensive behavioral/psychoeducational treatments for autism: Research needs and future directions. *Journal of Autism and Developmental Disorders*, 30(5), 373–378. <https://doi.org/10.1023/A:1005535120023>
- Stone, W. L., Coonrod, E. E., Turner, L. M., & Pozdol, S. L. (2004). Psychometric properties of the STAT for early autism screening. *Journal of Autism and Developmental Disorders*, 34(6), 691–701. <https://doi.org/10.1007/s10803-004-5289-8>

Young, R. L. (2007). *Autism Detection in Early Childhood: ADEC*. Australian Council of Educational Research, Camberwell, VIC, Australia.

Table 1

*Area Under the Curve (AUC) and 95% Confidence Intervals (CIs) for Individual Autism**Detection in Early Childhood (ADEC) Items*

ADEC Item	Nah et al. (2018)	Nevill et al. (2019)
	AUC [95% CI]	AUC [95% CI]
1. Response to name	.811 [.752, .864]	.743 [.653, .826]
2. Imitation	.698 [.627, .765]	.637 [.541, .732]
3. Ritualistic play	.626 [.562, .690]	.609 [.532, .687]
4. Joint attention/social referencing	.787 [.728, .842]	.683 [.588, .774]
5. Eye contact	.721 [.655, .785]	.645 [.544, .741]
6. Functional play	.753 [.687, .813]	.677 [.583, .768]
7. Pretend play	.717 [.649, .783]	.682 [.587, .772]
8. Reciprocity of smile	.783 [.718, .843]	.742 [.647, .830]
9. Reaction to common sounds	.700 [.641, .758]	.665 [.583, .746]
10. Gaze monitoring	.763 [.699, .823]	.825 [.751, .895]
11. Following verbal commands	.763 [.700, .822]	.738 [.645, .827]
12. Delayed language	.686 [.616, .757]	.687 [.600, .772]
13. Anticipation of social advances	.594 [.525, .659]	.640 [.547, .730]
14. Nestling	.641 [.580, .701]	.568 [.479, .657]
15. Use of gestures	.757 [.693, .818]	.629 [.531, .726]
16. Task switching	.722 [.655, .785]	.753 [.664, .836]

Note: Values in bold indicate the item's inclusion in the BADEC.

Table 2

*ADEC Items Ranked from Strongest to Weakest BADEC Contenders on the Basis of AUC**Comparisons in Nah et al. (2018) and Nevill et al. (2019)*

ADEC Item	Study	
	Nah et al. (2018)	Nevill et al. (2019)
Strongest BADEC Contenders		
Response to name	Highest AUC	No difference from highest AUC ($p = .136$)
Gaze monitoring	No difference from highest AUC ($p = .120$)	Highest AUC
Reciprocity of smile	No difference from highest AUC ($p = .449$)	No difference from highest AUC ($p = .129$)
Following verbal commands	No difference from highest AUC ($p = .194$)	No difference from highest AUC ($p = .129$)
Next Strongest Contenders		
Joint attention/social referencing	No difference from highest AUC ($p = .461$)	Lower than highest AUC ($p = .014$); No difference from other BADEC items ($ps > .220$)
Task switching	Lower than highest AUC ($p = .033$); No difference from other BADEC items ($ps > .106$)	No difference from highest AUC ($p = .147$)
Functional play	No difference from highest AUC ($p = .152$)	Lower than highest AUC ($p = .008$); No difference from other BADEC items ($ps > .209$)
Use of gestures	No difference from highest AUC ($p = .135$)	Lower than two highest AUCs ($ps < .036$); No difference from other BADEC items ($ps > .105$)
Cannot be Ruled Out		
Eye contact	Lower than highest AUC ($p = .021$); No difference from other BADEC items ($ps > .086$)	Lower than highest AUC ($p = .002$); No difference from other BADEC items ($ps > .086$)

Pretend play	Lower than highest AUC ($p = .022$); No difference from other BADEC items ($ps > .097$)	Lower than highest AUC ($p = .005$); No difference from other BADEC items ($ps > .210$)
Reaction to common sounds	Lower than two highest AUCs ($ps < .050$); No difference from other BADEC items ($ps > .060$)	Lower than highest AUC ($p = .004$); No difference from other BADEC items ($ps > .147$)
Imitation	Lower than two highest AUCs ($ps < .039$); No difference from other BADEC items ($ps > .062$)	Lower than two highest AUCs ($ps < .041$); No difference from other BADEC items ($ps > .091$)
Delayed language	Lower than five highest AUCs ($ps < .049$); No difference from other BADEC items ($ps > .076$)	Lower than highest AUC ($p = .008$); No difference from other BADEC items ($ps > .225$)
Anticipation of social advances	AUC lower than all BADEC items ($ps < .008$)	Lower than highest AUC ($p = .003$); No difference from other BADEC items ($ps > .068$)
Ruled Out		
Ritualistic play	AUC lower than all BADEC items ($ps < .005$)	AUC lower than all BADEC items ($ps < .038$)
Nestling	AUC lower than all BADEC items ($ps < .016$)	AUC lower than all BADEC items ($ps < .009$)

Note: For the Nah et al. (2018) dataset we have included in the “BADEC items” the two additional items (i.e., gaze monitoring and functional play) that were not part of Nah et al.’s BADEC but, on the basis of our significance testing, could have been used interchangeably with their BADEC items.

Supplemental Material

Table S1

R Code Showing the Calculation of AUC and 95% CIs, and the Comparison of AUC Between Two Items; Calculated from Items 1 and 2 (i.e., I01 and I02) of the ADEC from the Nah et al. (2018) Dataset.

```
## Load example dataset
ASDvsNTD <- read.csv("ExampleData.csv", header = TRUE)

## Calculate AUC and 95% CI for ADEC items 1 and 2
roc1 <- roc(ASDvsNTD$Diagnosis, ASDvsNTD$I01)
  print(roc1)
  ci.auc (roc1, method="bootstrap", boot.n=10000)
roc2 <- roc(ASDvsNTD$Diagnosis, ASDvsNTD$I02)
  print(roc2)
  ci.auc (roc2, method="bootstrap", boot.n=10000)

## compare AUC for ADEC items 1 and 2
roc.test(roc1, roc2, reuse.auc=FALSE, paired=TRUE, method="bootstrap", boot.n=10000)
```

Notes. The example dataset is available at <https://osf.io/utxrc/> ; for the diagnosis variable 1 = Non-

TD, 2 = ASD

Table S2

Nah et al (2018). Significance Values for Comparisons of Area Under the Curve (AUC) for the Individual Autism Detection in Early Childhood (ADEC) Items

ADEC Item	Autism Detection in Childhood (ADEC) Item															
	1: AUC = .811	2: AUC = .698	3: AUC = .626	4: AUC = .787	5: AUC = .721	6: AUC = .753	7: AUC = .717	8: AUC = .783	9: AUC = .700	10: AUC = .763	11: AUC = .763	12: AUC = .686	13: AUC = .594	14: AUC = .641	15: AUC = .757	16: AUC = .722
1: AUC = .811	-	.007	< .001	.461	.021	.152	.022	.449	.004	.120	.194	.004	< .001	< .001	.135	.033
2: AUC = .698			.126	.038	.562	.154	.650	.063	.901	.143	.089	.778	.036	.264	.174	.571
3: AUC = .626				< .001	.030	.004	.079	.001	.089	.003	.004	.226	.623	.852	.006	.020
4: AUC = .787					.087	.393	.098	.918	.049	.460	.620	.018	< .001	.001	.438	.107
5: AUC = .721						.454	.943	.087	.661	.320	.196	.374	.008	.116	.460	.987
6: AUC = .753							.334	.488	.239	.817	.727	.077	.001	.016	.985	.438
7: AUC = .717								.139	.767	.260	.208	.432	.015	.142	.274	.870
8: AUC = .783									.061	.607	.642	.033	< .001	.003	.488	.209
9: AUC = .700										.144	.198	.659	.015	.188	.185	.686
10: AUC = .763											.869	.048	< .001	.004	.919	.375
11: AUC = .763												.045	< .001	.005	.828	.340
12: AUC = .686													.090	.482	.099	.452
13: AUC = .594														.229	< .001	.010
14: AUC = .641															.007	.079
15: AUC = .757																.424
16: AUC = .722																-

Notes: Items highlighted in bold indicate they were included in the Brief Autism Detection in Early Childhood (BADEC). $N = 106$ in the ASD group; $N = 86$ in the non-ASD (i.e., non-typically developing) group

Table S3

Nevill et al (2019). Significance Values for Comparisons of Area Under the Curve (AUC) for the Individual Autism Detection in Early Childhood (ADEC) Items

ADEC Item	Autism Detection in Childhood (ADEC) Item																
	1: AUC = .743	2: AUC = .637	3: AUC = .609	4: AUC = .683	5: AUC = .645	6: AUC = .677	7: AUC = .682	8: AUC = .742	9: AUC = .665	10: AUC = .825	11: AUC = .738	12: AUC = .687	13: AUC = .640	14: AUC = .568	15: AUC = .629	16: AUC = .753	
1: AUC = .743	-	.125	.015	.272	.094	.313	.301	.993	.257	.136	.943	.339	.127	.010	.106	.852	
2: AUC = .637			.643	.463	.905	.541	.501	.092	.630	.001	.099	.375	.960	.264	.942	.040	
3: AUC = .609				.173	.532	.286	.174	.011	.276	< .001	.037	.150	.641	.489	.784	.011	
4: AUC = .683					.513	.940	.993	.294	.796	.014	.376	.947	.503	.081	.452	.221	
5: AUC = .645						.660	.552	.092	.743	.002	.160	.493	.948	.285	.717	.068	
6: AUC = .677							.933	.330	.858	.008	.335	.865	.590	.096	.506	.210	
7: AUC = .682								.317	.795	.005	.375	.933	.500	.055	.418	.211	
8: AUC = .742									.227	.129	.948	.289	.119	.005	.072	.846	
9: AUC = .665										.004	.250	.702	.681	.086	.522	.148	
10: AUC = .825											.129	.008	.003	< .001	.001	.147	
11: AUC = .738												.404	.073	.006	.108	.781	
12: AUC = .687													.462	.062	.328	.226	
13: AUC = .640														.120	.931	.069	
14: AUC = .568															.339	.002	
15: AUC = .629																.035	
16: AUC = .753																	-

Notes: Items highlighted in bold were included in the Brief Autism Detection in Early Childhood (BADEC). $N = 48$ in the ASD group; $N = 59$ in the non-ASD (i.e., ASD ruled out) group

Table S4

Nah et al (2018) Doubled. Significance Values for Comparisons of Area Under the Curve (AUC) for the Individual Autism Detection in Early Childhood (ADEC) Items

ADEC Item	Autism Detection in Childhood (ADEC) Item															
	1: AUC = .811	2: AUC = .698	3: AUC = .626	4: AUC = .787	5: AUC = .721	6: AUC = .753	7: AUC = .717	8: AUC = .783	9: AUC = .700	10: AUC = .763	11: AUC = .763	12: AUC = .686	13: AUC = .594	14: AUC = .641	15: AUC = .757	16: AUC = .722
1: AUC = .811	-	< .001	< .001	.300	.001	.042	.001	.286	< .001	.028	.068	< .001	< .001	< .001	.037	.003
2: AUC = .698			.030	.003	.404	.044	.516	.010	.857	.039	.015	.689	.003	.119	.058	.426
3: AUC = .626				< .001	.002	< .001	.013	< .001	.017	< .001	< .001	.087	.483	.789	< .001	.001
4: AUC = .787					.016	.226	.020	.885	.005	.287	.480	.001	< .001	< .001	.269	.022
5: AUC = .721						.294	.922	.015	.534	.162	.070	.208	< .001	.024	.296	.981
6: AUC = .753							.174	.326	.099	.743	.627	.011	< .001	.001	.979	.269
7: AUC = .717								.038	.677	.113	.078	.269	< .001	.036	.118	.816
8: AUC = .783									.006	.472	.506	.003	< .001	< .001	.323	.076
9: AUC = .700										.040	.071	.530	.001	.062	.062	.575
10: AUC = .763											.186	.006	< .001	< .001	.889	.204
11: AUC = .763												.005	< .001	< .001	.764	.175
12: AUC = .686													.015	.319	.019	.295
13: AUC = .594														.096	< .001	< .001
14: AUC = .641															< .001	.012
15: AUC = .757																.259
16: AUC = .722																-

Notes: Items highlighted in bold indicate they were included in the Brief Autism Detection in Early Childhood (BADEC). $N = 212$ in the ASD group; $N = 172$ in the non-ASD (i.e., non-typically developing) group

Table S5

Nevill et al (2019) Quadrupled. Significance Values for Comparisons of Area Under the Curve (AUC) for the Individual Autism Detection in Early Childhood (ADEC)

Items

ADEC Item	Autism Detection in Childhood (ADEC) Item																
	1: AUC = .743	2: AUC = .637	3: AUC = .609	4: AUC = .683	5: AUC = .645	6: AUC = .677	7: AUC = .682	8: AUC = .742	9: AUC = .665	10: AUC = .825	11: AUC = .738	12: AUC = .687	13: AUC = .640	14: AUC = .568	15: AUC = .629	16: AUC = .753	
1: AUC = .743	-	.002	< .001	.026	.001	.044	.041	.985	.025	.003	.886	.052	.002	< .001	.001	.714	
2: AUC = .637			.360	.146	.810	.209	.180	.001	.331	< .001	.001	.075	.919	.025	.885	< .001	
3: AUC = .609				.006	.210	.032	.007	< .001	.030	< .001	< .001	.004	.354	.165	.586	< .001	
4: AUC = .683					.197	.879	.987	.038	.597	< .001	.074	.893	.180	.001	.128	.013	
5: AUC = .645						.374	.236	.001	.514	< .001	.006	.170	.897	.034	.463	< .001	
6: AUC = .677							.866	.052	.716	< .001	.052	.738	.284	.001	.180	.011	
7: AUC = .682								.044	.602	< .001	.078	.865	.184	< .001	.104	.012	
8: AUC = .742									.017	.003	.895	.036	.002	< .001	< .001	.699	
9: AUC = .665										< .001	.019	.445	.426	.001	.201	.004	
10: AUC = .825											.003	< .001	< .001	< .001	< .001	.004	
11: AUC = .738												.094	< .001	< .001	.001	.576	
12: AUC = .687													.138	< .001	.051	.017	
13: AUC = .640														.002	.863	< .001	
14: AUC = .568															.057	< .001	
15: AUC = .629																< .001	
16: AUC = .753																	-

Notes: Items highlighted in bold were included in the Brief Autism Detection in Early Childhood (BADEC). $N = 192$ in the ASD group; $N = 236$ in the non-ASD (i.e., ASD ruled out) group

Table S6

Nah et al (2018). Correlations Between Autism Detection in Early Childhood (ADEC) Items

ADEC Item	Autism Detection in Childhood (ADEC) Item																
	1: AUC = .811	2: AUC = .698	3: AUC = .626	4: AUC = .787	5: AUC = .721	6: AUC = .753	7: AUC = .717	8: AUC = .783	9: AUC = .700	10: AUC = .763	11: AUC = .763	12: AUC = .686	13: AUC = .594	14: AUC = .641	15: AUC = .757	16: AUC = .722	
1: AUC = .811	-	.332**	.182*	.602**	.441**	.385**	.391**	.503**	.353**	.625**	.580**	.308**	.339**	.218**	.466**	.401**	
2: AUC = .698			.119	.311**	.260**	.463**	.394**	.216**	.248**	.294**	.412**	.347**	.151*	.073	.375**	.455**	
3: AUC = .626				.206**	.143*	.182*	.003	.171*	.196**	.165*	-.020	-.009	-.037	.188*	.046	.327**	
4: AUC = .787					.488**	.447**	.373**	.487**	.250**	.626**	.481**	.322**	.188**	.198**	.397**	.442**	
5: AUC = .721						.379**	.281**	.451**	.163*	.472**	.436**	.356**	.175*	.117	.353**	.265**	
6: AUC = .753							.515**	.376**	.314**	.446**	.430**	.486**	.065	.168*	.522**	.400**	
7: AUC = .717								.328**	.068	.409**	.453**	.410**	.112	.130	.511**	.274**	
8: AUC = .783									.293**	.489**	.427**	.231**	.225**	.163*	.391**	.241**	
9: AUC = .700										.308**	.220**	.219**	.088	.098	.222**	.241**	
10: AUC = .763											.634**	.476**	.257**	.223**	.432**	.313**	
11: AUC = .763												.380**	.283**	.160*	.456**	.264**	
12: AUC = .686													.123	.067	.402**	.336**	
13: AUC = .594														.288**	.298**	.168*	
14: AUC = .641															.239**	.208**	
15: AUC = .757																.353**	
16: AUC = .722																	-

** Correlation significant < .01; * correlation significant < .05

Table S7

Nevill et al (2019). Correlations Between Autism Detection in Early Childhood (ADEC) Items

ADEC Item	Autism Detection in Childhood (ADEC) Item																
	1: AUC = .743	2: AUC = .637	3: AUC = .609	4: AUC = .683	5: AUC = .645	6: AUC = .677	7: AUC = .682	8: AUC = .742	9: AUC = .665	10: AUC = .825	11: AUC = .738	12: AUC = .687	13: AUC = .640	14: AUC = .568	15: AUC = .629	16: AUC = .753	
1: AUC = .743	-	.040	.271**	.399**	.321**	.161	.283**	.381**	-.019	.402**	.239*	.290**	.069	-.069	.060	.316**	
2: AUC = .637			.082	.200*	.120	.147	.092	.210*	.144	.237*	.259**	.350**	.176	.044	.241*	.313**	
3: AUC = .609				.347**	.265**	.001	.292**	.276**	.190	.388**	.066	.217*	-.045	.012	.083	.159	
4: AUC = .683					.294**	-.121	.190*	.330**	.048	.290**	.240*	.107	.182	.012	.119	.276**	
5: AUC = .645						-.013	.246*	.326**	.147	.372**	.081	.179	-.001	-.117	.222*	.344**	
6: AUC = .677							.292**	.090	.090	.301**	.273**	.281**	.058	.021	.231*	.269**	
7: AUC = .682								.267**	.120	.411**	.263**	.385**	.156	.166	.272**	.331**	
8: AUC = .742									.081	.385**	.337**	.416**	.152	.047	.249*	.341**	
9: AUC = .665										.190	.127	.232*	.087	.152	.138	.169	
10: AUC = .825											.330**	.343**	.133	-.002	.286**	.452**	
11: AUC = .738												.261**	.362**	.112	.198*	.431**	
12: AUC = .687													.125	-.024	.327**	.328**	
13: AUC = .640														.507**	.103	.182	
14: AUC = .568															-.093	.108	
15: AUC = .629																.328**	
16: AUC = .753																	-

** Correlation significant < .01; * correlation significant < .05

Table S8

Nah et al (2018) Highest Possible Between-item Correlations. Correlations Between Autism Detection in Early Childhood (ADEC) Items

ADEC Item	Autism Detection in Childhood (ADEC) Item																
	1: AUC = .811	2: AUC = .698	3: AUC = .626	4: AUC = .787	5: AUC = .721	6: AUC = .753	7: AUC = .717	8: AUC = .783	9: AUC = .700	10: AUC = .763	11: AUC = .763	12: AUC = .686	13: AUC = .594	14: AUC = .641	15: AUC = .757	16: AUC = .722	
1: AUC = .811	-	.866**	.710**	.920**	.849**	.892**	.851**	.938**	.762**	.919**	.919**	.868**	.671**	.693**	.894**	.856**	
2: AUC = .698			.808**	.885**	.909**	.936**	.811**	.850**	.801**	.903**	.930**	.883**	.829**	.805**	.896**	.898**	
3: AUC = .626				.777**	.823**	.823**	.552**	.664**	.887**	.751**	.810**	.696**	.898**	.930**	.714**	.760**	
4: AUC = .787					.885**	.949**	.787**	.882**	.824**	.962**	.952**	.883**	.724**	.755**	.927**	.868**	
5: AUC = .721						.911**	.744**	.825**	.833**	.880**	.910**	.845**	.843**	.819**	.892**	.868**	
6: AUC = .753							.777**	.872**	.824**	.950**	.966**	.875**	.817**	.818**	.930**	.874**	
7: AUC = .717								.876**	.525**	.795**	.810**	.895**	.538**	.510**	.798**	.828**	
8: AUC = .783									.717**	.905**	.880**	.922**	.640**	.651**	.910**	.878**	
9: AUC = .700										.794**	.813**	.663**	.856**	.923**	.757**	.772**	
10: AUC = .763											.937**	.905**	.725**	.731**	.963**	.836**	
11: AUC = .763												.881**	.772**	.788**	.917**	.890**	
12: AUC = .686													.674**	.646**	.919**	.871**	
13: AUC = .594														.908**	.688**	.762**	
14: AUC = .641															.695**	.764**	
15: AUC = .757																.817**	
16: AUC = .722																	-

**Correlation significant < .01

Table S9

Nevill et al (2019) Highest Possible Between-item Correlations. Correlations Between Autism Detection in Early Childhood (ADEC) Items

ADEC Item	Autism Detection in Childhood (ADEC) Item																
	1: AUC = .743	2: AUC = .637	3: AUC = .609	4: AUC = .683	5: AUC = .645	6: AUC = .677	7: AUC = .682	8: AUC = .742	9: AUC = .665	10: AUC = .825	11: AUC = .738	12: AUC = .687	13: AUC = .640	14: AUC = .568	15: AUC = .629	16: AUC = .753	
1: AUC = .743	-	.666**	.686**	.901**	.884**	.876**	.755**	.873**	.824**	.914**	.874**	.641**	.892**	.744**	.716**	.843**	
2: AUC = .637			.369**	.662**	.757**	.663**	.876**	.838**	.452**	.596**	.841**	.923**	.668**	.486**	.892**	.804**	
3: AUC = .609				.762**	.652**	.770**	.441**	.553**	.899**	.688**	.547**	.355**	.575**	.782**	.364**	.550**	
4: AUC = .683					.878**	.943**	.759**	.834**	.833**	.787**	.838**	.637**	.864**	.849**	.716**	.801**	
5: AUC = .645						.872**	.822**	.904**	.767**	.763**	.905**	.728**	.882**	.805**	.809**	.866**	
6: AUC = .677							.775**	.823**	.843**	.777**	.828**	.638**	.866**	.796**	.713**	.787**	
7: AUC = .682								.862**	.539**	.644**	.865**	.896**	.797**	.581**	.880**	.882**	
8: AUC = .742									.677**	.810**	.988**	.844**	.854**	.676**	.838**	.939**	
9: AUC = .665										.827**	.669**	.434**	.704**	.798**	.467**	.673**	
10: AUC = .825											.813**	.586**	.731**	.596**	.612**	.783**	
11: AUC = .738												.846**	.843**	.667**	.851**	.927**	
12: AUC = .687													.642**	.468**	.870**	.854**	
13: AUC = .640														.758**	.735**	.837**	
14: AUC = .568															.516**	.667**	
15: AUC = .629																.819**	
16: AUC = .753																	-

** Correlation significant < .01

Table S10

Nah et al (2018) Highest Possible Between-item Correlations. Significance Values for Comparisons of AUC for the Individual Autism Detection in Early Childhood

(ADEC) Items

ADEC Item	Autism Detection in Childhood (ADEC) Item															
	1: AUC = .811	2: AUC = .698	3: AUC = .626	4: AUC = .787	5: AUC = .721	6: AUC = .753	7: AUC = .717	8: AUC = .783	9: AUC = .700	10: AUC = .763	11: AUC = .763	12: AUC = .686	13: AUC = .594	14: AUC = .641	15: AUC = .757	16: AUC = .722
1: AUC = .811	-	< .001	< .001	.154	< .001	< .001	< .001	.023	< .001	.007	< .001	< .001	< .001	< .001	.001	< .001
2: AUC = .698			.002	< .001	.333	< .001	.434	< .001	.708	< .001	< .001	.439	< .001	.028	< .001	.272
3: AUC = .626				< .001	< .001	< .001	.004	< .001	< .001	< .001	< .001	.013	.005	.142	< .001	< .001
4: AUC = .787					< .001	.001	.001	.835	< .001	.029	.001	< .001	< .001	< .001	.006	< .001
5: AUC = .721						.009	1	.001	.290	.001	.007	.111	< .001	.003	.007	.760
6: AUC = .753								.110	.044	.006	.106	.761	< .001	< .001	< .001	.952
7: AUC = .717									.001	.431	.025	.072	.121	< .001	.045	.138
8: AUC = .783										< .001	.226	.055	< .001	< .001	< .001	.039
9: AUC = .700											< .001	.004	.938	< .001	.002	.005
10: AUC = .763												.311	< .001	< .001	< .001	.153
11: AUC = .763													< .001	< .001	< .001	.860
12: AUC = .686														< .001	.160	< .001
13: AUC = .594															< .001	< .001
14: AUC = .641																< .001
15: AUC = .757																
16: AUC = .722																

Notes: Items highlighted in bold indicate they were included in the Brief Autism Detection in Early Childhood (BADEC). $N = 106$ in the ASD group; $N = 86$ in the non-ASD (i.e., non-typically developing) group

Table S11

Nevill et al (2019) Highest Possible Between-item Correlations. Significance Values for Comparisons of AUC for the Individual Autism Detection in Early Childhood

(ADEC) Items

ADEC Item	Autism Detection in Childhood (ADEC) Item																
	1: AUC = .743	2: AUC = .637	3: AUC = .609	4: AUC = .683	5: AUC = .645	6: AUC = .677	7: AUC = .682	8: AUC = .742	9: AUC = .665	10: AUC = .825	11: AUC = .738	12: AUC = .687	13: AUC = .640	14: AUC = .568	15: AUC = .629	16: AUC = .753	
1: AUC = .743	-	.002	< .001	.001	< .001	.002	.056	.983	.008	< .001	.849	.147	< .001	< .001	< .001	.699	
2: AUC = .637			.568	.183	.804	.244	.088	< .001	.536	< .001	< .001	.003	.934	.131	.814	< .001	
3: AUC = .609				.047	.355	.068	.111	.001	.024	< .001	.002	.106	.438	.125	.557	< .001	
4: AUC = .683					.083	.718	.985	.022	.594	< .001	.027	.910	.086	< .001	.084	.012	
5: AUC = .645						.104	.179	< .001	.508	< .001	< .001	.240	.846	.015	.709	< .001	
6: AUC = .677							.845	.017	.722	< .001	.020	.793	.105	.001	.131	.009	
7: AUC = .682								.031	.690	< .001	.044	.854	.148	.004	.036	.005	
8: AUC = .742									.033	.002	.696	.044	< .001	< .001	< .001	.542	
9: AUC = .665										< .001	.047	.644	.433	< .001	.493	.011	
10: AUC = .825											.001	< .001	< .001	< .001	< .001	.022	
11: AUC = .738												.065	< .001	< .001	< .001	.470	
12: AUC = .687													.238	.009	.030	.017	
13: AUC = .640														.023	.863	< .001	
14: AUC = .568															.125	< .001	
15: AUC = .629																< .001	
16: AUC = .753																	-

Notes: Items highlighted in bold were included in the Brief Autism Detection in Early Childhood (BADEC). $N = 48$ in the ASD group; $N = 59$ in the non-ASD (i.e., ASD ruled out) group