Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v12n2p320

Preliminary evidence of the structural validity and measurement invariance of the Quantitative-CHecklist for Autism in toddler (Q-CHAT) on Italian unselected children

By Lecciso et al.

Published: 14 October 2019

Electronic Journal of Applied Statistical Analysis
Vol. 12, Issue 02, October 2019, 320-340
DOI: 10.1285/i20705948v12n2p320

Preliminary evidence of the structural validity and measurement invariance of the Quantitative-CHecklist for Autism in toddler (Q-CHAT) on Italian unselected children

Flavia Lecciso*[a,c], Annalisa Levante[a,c], Fulvio Signore[a], and Serena Petrocchi[b,c]

[a]*Department of History, Society and Human Studies, University of Salento, Italy*
[b]*Institute of Communication and Health, Università della Svizzera Italiana, Switzerland*
[c]*Lab of Applied Psychology and Intervention, University of Salento, Italy*

Published: 14 October 2019

The Quantitative-CHecklist for Autism in toddler (Q-CHAT) is a screening measure developed to detect the early signs of risk of Autism Spectrum Disorders (ASD) in children from 18-24 months of life. After a critical analysis of the validation studies that have analyzed the Q-CHAT, the present paper aimed at testing the structural validity of the questionnaire through Confirmatory Factor Analysis and determining whether it is invariant across gender. Furthermore, the study examined the internal validity of the measure, calculated the risk cut-off score, and evaluated gender differences. Five hundred forty-five questionnaires were completed by parents of children from unselected population (age M = 19.9 months; *sd* = 1 month; age range = 18-21; 254 girls). Findings gave tolerable demonstration of the expected three-factor structure and measurement invariance across gender. Gender differences were also found as well as correlations among factors as a demonstration of internal validity. The risk cut-off of 43 was calculated based on the 95 th percentile of the distribution. In conclusion, the present study demonstrated preliminary findings of the validity of the Q-CHAT for Italian children. However, our findings suggested also that three items need to

*Corresponding author: flavia.lecciso@unisalento.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

be reformulated because they show several psychometric problems and, after that, the factorial validity, measurement invariance, and internal validity should be re-tested. Further longitudinal studies are also needed to examine the validity criterion of the Q-CHAT to test the cut-off score.

keywords: Q-CHAT, Quantitative-CHecklist for Autism in Toddler, autism spectrum disorders, confirmatory factor analysis, measurement invariance, internal validity.

1 Introduction

Autism Spectrum Disorders (henceforth, ASD) are a group of neurodevolopmental conditions characterized by social communicative deficits and repetitive/stereotyped behaviors (APA, 2013). Key behaviors detecting early warning signs of ASD have been found in children under 24 months of life (e.g. Robins et al., 2001; Zwaigenbaum et al., 2015) and several screening measures have been developed with the purpose to identify those early signs, with different features in term of validity (see Levante et al., 2019, a,b). The main purpose of these measures is to develop a systematic screening procedure to early detect children at risk for ASD on general population.
The Quantitative-CHecklist for Autism in Toddler (Q-CHAT; Allison et al., 2008), which is a parent-reported questionnaire during regular well-baby exams when child aged 18-24 months of life, is one of the most known. This measure was developed by Allison and colleagues (2008) and this first validation study was followed by others on general population (Magiati et al., 2015; Ruta et al., 2019).

The Q-CHAT is the revised version of the first borne CHecklist for Autism in Toddler (CHAT; Baron-Cohen et al., 1992) and the subsequent Modified-CHecklist for Autism in Toddler (M-CHAT; Robins et al., 2001), which focused on joint attention, social and pretend play, social interest, and declarative pointing. The rationale is that when these psychological behaviors are absent, then the child is at risk for ASD (Baron-Cohen et al., 1992). The main limitation of these two versions of the questionnaire was the fact that the items are rated on a dichotomous scale, in which the targeted behaviors are evaluated as present or absent. This scale of measure has been considered too stringent and conservative (Allison et al., 2008). Moreover, the sensitivity of the CHAT, which is the degree of which a tool identifies as positive a case when a given condition is actually present, was low (i.e., 38%; see Baron-Cohen et al., 2000).

Given these limitations, a new version called Q-CHAT was developed. The main innovation of the Q-CHAT was the change on the response format that became on a 5-point Likert-scale. Indeed, Allison and colleagues (2008) suggested that evaluating the frequency of atypically behaviors, instead of their presence or absence, allows the identification even of the mildest case of ASD. Moreover, compared to the first two versions of the same measure, in the Q-CHAT several items were added tapping language development, repetitive behaviors, and social communication behaviors, according to the main diagnostic manuals of that time (ICD-10, WHO, 1993; DSM-IV-TR, APA, 1994). The final version of the measure includes 25 item.

2 A critical analysis on the previous studies of the Q-CHAT

To our best knowledge, there are three published validation studies on the Q-CHAT (Allison et al., 2008; Magiati et al., 2015; Ruta et al., 2019). The first one by Allison and colleagues (2008) examined the Q-CHAT score distribution in a group of unselected children and in a group of children with ASD. The authors reported a near-normal data distribution in the unselected group (n = 660; M = 21.2 months; $sd = 2.1$ months) and a normal distribution in the ASD group (n = 160; M = 44.5 months; $sd = 10.2$ months). The internal consistency of the Q-CHAT was moderate ($\alpha = .67$) for the unselected group and acceptable ($\alpha = .83$) for the ASD group. The test-retest reliability for the unselected group showed a strong correlation between the two-time points (with a mean of days of time interval of 38). The authors found that five items had the highest number of missing data (i.e., items 3, 10, 15, 22, and 23) and for those items they applied a conservative approach imputing a 0 to each of them (corresponding to the low frequency of the atypically behavior). Furthermore, the authors found that boys scored higher than girls in both groups. In the unselected group, the age and the Q-CHAT score were not associated, whereas, in the ASD group of children, the total score of the Q-CHAT decreased with age. Finally, in the group of children with ASD, the total score was significantly higher than that of the unselected group.

Magiati and colleagues (2015) designed a study to examine the psychometric properties of the Q-CHAT in a Singaporean unselected group of children at 18 ($n = 368$) and 24 ($n = 396$) months of life. Their findings showed a normal distribution of the Q-CHAT scores in both groups. Details on missing data were not available in the paper with the exception of the imputation method applied, i.e., the same conservative approach suggested by Allison et al. (2008). The authors carried out an Exploratory Factor Analysis (EFA) applying the Principal Axis Factoring method and the oblique direct oblimin rotation. The 3-factor structure excluded the items 3, 13, and 14 since their factor loadings were under the threshold of .30. The final structure included the "Social-Communicative autistic traits" (Factor 1: item 1, 2, 5, 6, 9, 10, 12, 15, 19, and 21), the "Non-social/Behavioral autistic traits" (Factor 2: item 7, 11, 16, 20, 22, 23, 24, and 25), and the "Speech/Language items" (Factor 3: item 4, 8, 17, and 18).

As regards to the internal consistency, findings showed from moderate to low evidence of reliability in both groups of children (18 - 24 months of life). The Q-CHAT total score reached a Cronbach's alpha of .53 and .60 respectively. The Social-Communicative autistic traits factor showed a .76 at 18 months and a .75 at 24 months of life. The Non-social/Behavioral autistic traits factor reached a .69 at 18 months and a .71 at 24 months and, finally, the Speech/Language factor a .63 both at 18 and at 24 months. On the same way, even results from the item-total correlations showed mixed evidence of reliability (see Magiati et al., 2015 for details).

Finally, Magiati and colleagues (2015) demonstrated the convergent validity of the Q-CHAT, through correlations between the total score, the Social-Communicative autistic trait, and the Non-social/Behavioral autistic trait with the Pervasive Developmental Disorder scale assessed by the Child Behavior CheckList (Achenbach and Rescorla, 2000).

Ruta et al. (2019) administered the Q-CHAT for the first time on an Italian unselected population (n = 2400; M = 23.2; *sd* = 5.2) with the main aim to define a systematic screening procedure for ASD within routine pediatric developmental surveillance. Italian data were normally distributed and the total score was lower than the UK (Allison et al., 2008) and Singaporean ones (Magiati et al., 2015). Full details of missing data for every single item were not available, but the authors reported the overall percentage of the full questionnaires (3.1%) excluded from the analyses since seven or more items were missing. Significant difference by gender was found, with boys showing higher score than girl. As done by Magiati and colleagues (2015), the authors examined the item-total correlations and results showed mixed evidence (see Ruta et al., 2019 for details).

The authors carried out an EFA following the same procedure reported in Magiati and colleagues (2015) study. A factor structure was extracted from the Italian data, tapping three factors called "Non-Social/Behavioral Autistic traits" (Factor 1: item 3, 7, 11,12, 16, 20, 22, 23, 24, and 25), "Speech/Language" (Factor 2: item 1, 4, 8, 15, 17, and 18), and "Joint attention/non-verbal communication" (Factor 3: item 5, 6, 10, 12, and 19). The items 2, 9, 13, 14, and 21 were excluded from the analysis since their factor loadings were under the threshold of .30.

Magiati and colleagues (2015) and Ruta and colleagues (2019) factor structures presented some similarities and several differences. The Non-social/Behavioral autistic traits factor, extracted by Ruta and colleagues (2019), corresponds to the Non-social/ Behavioral autistic traits factor, extracted by Magiati and colleagues (2015), with the addition to items 3 and 12. The Speech/ Language factor, extracted by Ruta and colleagues (2019), corresponds to the Speech/Language factor, extracted by Magiati and colleagues (2015), with the addition to items 1 and 15. Finally, the Joint attention/non-verbal communication factor, extracted by Ruta and colleagues (2019), partially corresponds to the Social-Communicative autistic traits factor, extracted by Magiati and colleagues (2015), with the exception of the items 2, 9, and 21. In both structures, the item 12 loaded for two factors with opposite signs.

The three validation studies gave demonstrations of construct validity (Allison et al., 2008; Magiati et al., 2015), concurrent validity (Allison et al., 2008), and structural validity via exploratory factor analysis (Magiati et al., 2015; Ruta et al., 2019) of the Q-CHAT. Although they demonstrated the methodological quality of the Q-CHAT, they have some limitations. Specifically, none of them demonstrated the structural validity of the Q-CHAT through Confirmatory Factor Analysis (CFA) and none of the previous studies on the Q-CHAT analyzed the measurement invariance across gender. The measurement invariance is an essential and strategic analysis to perform cross-cultural study (see also Petrocchi et al., 2017) and its aim is to verify whether a construct is measured in a similar way across some specified groups is to verify whether a construct is measured in a similar way across some specified groups. Any violation of this assumption precludes meaningful interpretation of measurement and comparison. Since the ratio of the ASD for boys and girls is 4:1 (APA, 2013), it is important to establish whether the structural validity of the Q-CHAT is gender invariant. Finally, none of them calculated the cut-off score to detect children at risk in the general population at 18 months of life.

The present paper aimed at overcomes the above mentioned limitations testing:

a) a) The Q-CHAT structural validity carrying out, for the first time, a CFA based on the exploratory factor structure found by Ruta et al. (2019) on an Italian sample;

b) Whether there was a measurement invariance across gender;

c) The internal validity of the Q-CHAT;

d) And calculating the cut-off scores to detect children at risk.

The CFA and the measurement invariance, in particular, represent the added value of this study in comparison with Ruta et al. and colleagues' one (2019) which is on the same population. The present study applied the Q-CHAT on general population. Having a tool to be easily employed in a systematic population based screening for ASD would be useful for the early detection of the syndrome and the early start of appropriate treatments (Anderson et al., 2014; Daniels and Mandell, 2014; Dawson et al., 2010; Perry et al., 2008; Renty and Roeyers, 2006; Robins and Dumont-Mathieu, 2006; Sallows and Graupner, 2005; Sutera et al., 2007). Studies focused on unselected population are useful to improve the validity of the measure, to discriminate between typical and atypical developmental situations, and to the identification from sever to mild autistic traits.

3 Methods

3.1 Procedure

The current study uses data from wave 2 of a three-wave longitudinal study. The Ethical Committee of the Local Public Health Service gave its approval for the research (n^o 528/8 March 2017). Date were collected in an urban area of a Southern Region of Italy. One hundred fifteen pediatricians of the local public health service received via mail a description of the research project and a request to collaborate with it. 55.6% (n = 64) of them participated in the research and received instructions for recruitment of participants. All families treated by those pediatricians with a child born between February and September of 2016 were invited to participate in the study (n = 800). All families were contacted by their pediatricians and received a cover letter with the research project description and the informed consent. Parents agreed to participate completed alone questionnaire during well-child check-ups in a quiet room of the pediatrician' office.

3.2 Participants

698 questionnaires were delivered and 545 questionnaires have been returned completed, with a response rate of 78.08%. The questionnaires were completed by mothers for 71.2% of the cases, by father for 5.7%, and by both parents for the 22.8%. Two hundred eighty-eight questionnaires were completed for boys and 254 for girls (M = 19.9 months; *sd* = 1 month; range = 18-21). The children's mean weight at birth was 3.30 kg (*sd* = .56); 398 toddlers were born at term (after the 38th gestation weeks) and 121 were born pre-term. Two hundred and twenty-three toddlers were first-born and 234 were second-born or more.

Parental variables examined were age, educational level and marital status. Mothers' mean of age was 34 (*sd* = 5.5 years) and their educational level was low (up to 8 years of education) for 22%, intermediate (up to 13 years of education) for 45.3%, and high (18 or more years of educational) for 28.8%. Father's mean of age was 37.6 (*sd* = 6.5 years) and their educational level was low (up to 8 years of education) for 32.8%, intermediate (up to 13 years of education) for 40.7%, and high (18 or more years of educational) for 19.1%. The majority of the parents were married (91.7%), whereas only 7.3% were single or divorced.
Table 1 summarizes the socio-demo characteristics of the sample and the results of the t-tests by children's gender. The only significant difference was found on children's birth weight.

Table 1: Descriptive of the sample divided by children gender

Variables	**Male**	**Female**	**t-test**
Children's age M (DS)	19.9(1.7)	19.9(1)	t(538) = -.039; p >.05
Number of term	Term = 213;	Term = 183;	t(514) = -.247; p >.05
and preterm pregnancy	preterm = 63	preterm = 57	
Mean (and DS) of birth weight	3.37 (.52)	3.21 (.58)	t(519) = 3.437; p =.001
Level of parental education	High = 85;	High = 71;	t(519) = .759; p >.05
	intermediate = 134;	intermediate = 112;	
	low = 59;	low = 60;	

Note: high = 18 or more years of educational; intermediate = up to 13 years of education; low = up to 8 years of education.

3.3 Measures

The Q-CHAT is a parent-reported questionnaire developed to detect the autistic traits frequency in toddlers aged 18–24 months. The tool includes 25 items (13 of them are reverse scored) with response options on a 5-point Likert scale (0–4). The Q-CHAT total score is calculated as a sum of all items and its theoretical range varies from 0 to 100 with higher scores corresponding to higher risk of ASD. For this research, parents completed the Italian version of the Q-CHAT (Levante et al., 2017). For the translation of the questionnaire, a 2-step back-translation procedure was followed: the English version of the Q-CHAT was first translated into Italian by a native speaker; an English native speaker who is fluent also in Italian back-translated the questionnaire in English. The two English versions did not show any differences, and then the Italian version of the questionnaire was send by e-mail to the original authors who published it online. Children's (i.e., age, gender, birthweight, birth order) and parents' (i.e., age, marital status, educational level, and ethnicity) socio-demo data were collected.

3.4 Statistical analyses

Preliminary analyses, inter-item correlations, and reliability analysis through Omega were conducted using *SPSS v.25* and *JASP v.0.10.1* for Mac. Confirmatory Factor Analysis (CFA) and invariance (Crisci and D'Ambra, 2012) measurement across gender were conducted using *RStudio v.1.2.1335* applying the *Lavaan* (Rosseel, 2012) and *semTools* (Pornprasertmanit et al., 2015) packages. For the CFA, as goodness-of-fit indices, we assessed the chi-square, the root mean square error of approximation (RMSEA), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Standardized Root Mean Square Residual (SRMR). The chi-square should be non-significant although is influenced by the sample size and often is actually significant even if the data well fit the model (Brannick, 1995; Kelloway, 1995; Ciavolino, 2012). Acceptable values for RMSEA and SRMR are $< .06$, for CFI and TLI $> .95$ (i.e. Byrne, 2010). The CFA tested the validity of the factor structure explored by Ruta and colleagues (2019) since it was the only one established on an Italian sample of children.

Invariance was tested by using:

1. *Configural* invariance, in which the aim is to check if the organization of the different dimensions is the same in the groups. Configural invariance is based on the overall fit of the model. No constraints are imposed.

2. *Metric* invariance, where each item should contribute in a similar degree to a latent construct. The metric model, or weak invariance, is evaluated by constraining factor loadings.

3. *Scalar* invariance, which measures strong invariance in case of equivalent item intercept in the two groups of male and female. In this case, the we constrained factor loadings and intercepts.

4. *Residual* invariance, or equivalence of residual variance and measurement error. Finally, the constrains are related to factor loadings, intercepts, residuals and means of latent factors. For metric, scalar and residual models, invariance is

evaluated by comparing the fit of two nested models that are identical except for a target set of restrictions in one (Putnick and Bornstein, 2016).

The fit of the configural invariance model was compared with the fit of the nested models using the difference between Chi-squares (i.e., $\Delta\chi^2$). A non-significant $\Delta\chi^2$ is evidence of invariance. However, as the Chi-square, even the ΔChi-square depends on sample size (Brannick, 1995; Kelloway, 1995); therefore, we used the ΔCFI (i.e., change in CFI) paired with ΔRMSEA (i.e., change in RMSEA) as suggested by Cheung and colleagues (2002) and Chen et al. (2008): ΔCFI values smaller than or equal to .01 and ΔRMSEA smaller than .015 are considered evidence of invariance.

4 Results

4.1 Preliminary analyses

The preliminary analyses showed that the items had less than 5% of missing data. We decided not to replace the missing data both because the 1.1% was under the warning threshold Schafer and Graham (2002) and because the R package applied has a strong algorithm applicable even in case of missing data. The Q-CHAT total score is normally distributed (see Figures 1 and 2) with skewness = .185 and kurtosis = .028.

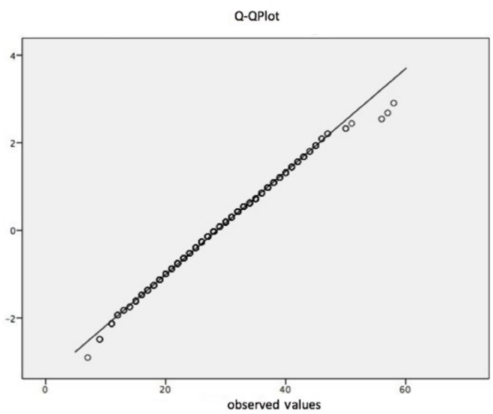

Figure 1: Q-Q plot of the Q-CHAT total score

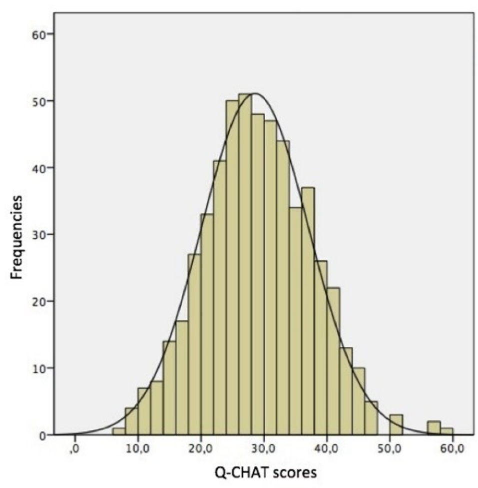

Figure 2: Distribution of the Q-CHAT total score

Independent sample t-tests on the total score and factor-score by children's gender showed significant differences on the total score ($t(539.908) = 2.273$; $p = .02$) and the speech and language factor ($t(521.013) = 2.013$; $p = .045$). As found in other research (e.g. Fabio et al., 2011), boys reached higher scores (total score: M = 29.35; $sd = 9$; speech and language factor: M = 7.40; $sd = 2$) than girls (total score: M = 27.72; $sd = 7.8$; speech and language factor: M = 7.09; $sd = 1.6$).

4.2 Confirmatory Structural Validity of the Q-CHAT

Figure 3 shows the theoretical model tested based on the Italian exploratory factor structure Ruta et al. (2019), which was described in the introduction section. The item 12, in the Ruta and colleagues (2019) model, loaded for two factors: the Non-social/Behavioral autistic traits (factor 1), with a positive sign, and the Joint attention/Non-verbal communication (factor 3), with a negative sign. Whereas the item 18 loaded for the Speech/Language factor with a minus sign (see Ruta et al., 2019).

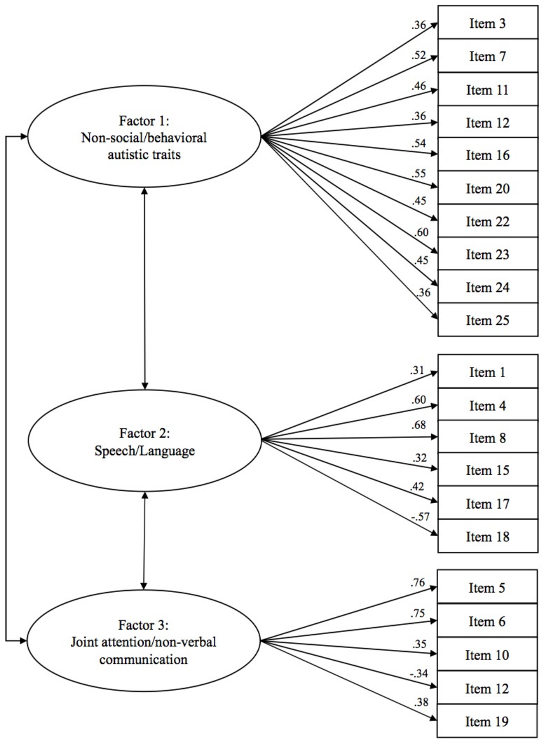

Figure 3: Theoretical model of the tested CFA

In the first CFA, we included both the items 12 and 18 to test how they load on the expected factors and to examine the same factor structure tested by Ruta and colleagues (2019). The resulting CFA showed a significant χ^2 (166) = 460.34, $p < .0001$, CFI = .84, TLI = .81, RMSEA = .059, SRMR = .06. The item 12 loaded for the Non-social/Behavioral autistic traits factor (β = .25) and the Joint attention/Non-verbal communication factor (β = -.22). We opted to delete this item from the subsequent analyses for three reasons. First, higher scores on the item 12 correspond to higher risk of ASD. Therefore, this item should not load any factor with a minus sign. Second, the two factor loadings are quite similar and, third, the values of the factor loadings are < .30. Similarly, the item 15 showed a factor loading of .23 for the speech and language factor. Since the value is lower than .30, this item was deleted from the subsequent analysis.

A second CFA was tested without the items 12 and 15. The CFA showed a significant $\chi 2$ (132) = 393.12, p < .0001, CFI = .85, TLI = .83, RMSEA = .062, SRMR = .06. The item 18 loaded with a minus sign for the Speech/Language factor. Based on the modification indices, the item was forced to load the Non-social/Behavioral autistic traits factor. The resulting CFA showed slightly worse fit indices χ^2 (149) = 450.57, $p < .0001$, CFI = .83, TLI = .81, RMSEA = .063, SRMR = .069. Thus, the item 18 was finally included in the Speech/Language factor. Figure 4 shows the final model derived from the CFA with the results of the analysis.

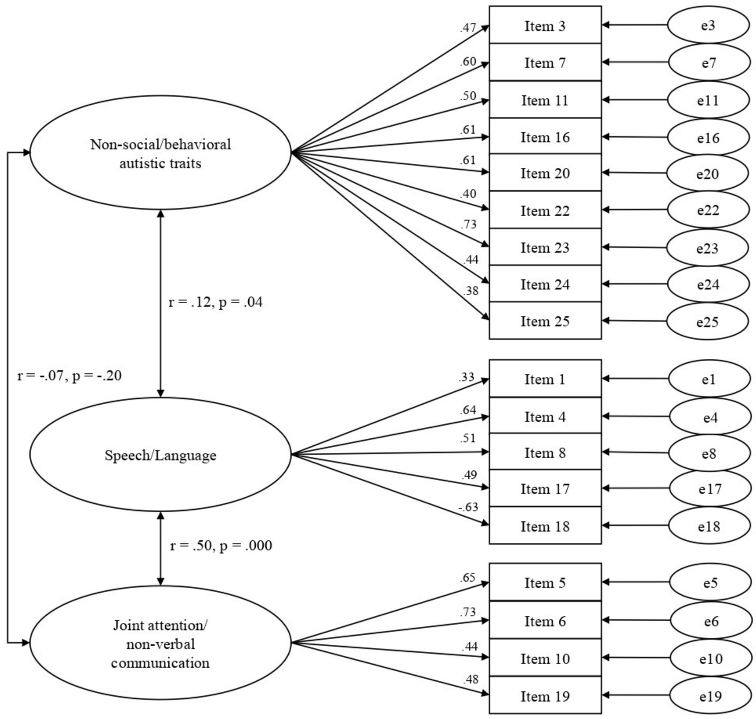

Figure 4: CFA on Italian Sample. All the factor loadings are significant at p < .05

McDonald's Omega was calculated for all the three factors showing from moderate to low reliability (Ω factor 1 = .78; Ω factor 2 = .33; Ω factor 3 = .67).

4.3 Measurement invariance

The final factor structure was tested for measurement invariance across gender. Results of the analyses showed a tolerable configural invariance: $\chi 2$ is significant (p < .001), CFI = .83 and TLI = .81 are not good while RMSEA = .063, SRMR = .069, GFI = .92 and AGFI = .897 have an acceptable fit. By considering the models divided by boys and girls, we found that the best fit is for girls, where $\chi 2$ (132) = 257.58, $p < .0001$, CFI = .84, TLI = .815, RMSEA = .063, and SRMR = .069. For the boys, χ^2 (132) = 309.638, $p <.001$, CFI = .822, TLI = .794, RMSEA = .070, and SRMR = .073. Metric invariance and scalar invariance were met, whereas residual invariance was not. Table 2 shows the results of the measurement invariance.

4.4 Internal validity of the Q-CHAT

The findings highlighted a significant and positive correlation between the speech and language factor and the Non-social/Behavioral autistic traits factor (r = .17, p = .000) and between the Joint attention/Non-verbal communication factor and the Speech/ Language factor (r = .16, p = .000).

Table 2: Measurement invariance across gender

Model	χ^2	$\Delta\chi^2$	Df	Δ Df	Δ p	CFI	Δ CFI	RMSEA	Δ RMSEA
Configural invariance	567.22	-	264	-	< .001	.830	-	.067	-
Metric invariance	599.36	32.146	279	15	.006	.820	.010	.067	.000
Scalar invariance	744.64	120.098	312	18	.000	.757	.057	.073	.007
Residual invariance	751.22	6.586	315	3	.086	.755	.002	.073	.000

4.5 Cut-Offs for the Detection of the ASD Risk

As we reported above, no previous studies identified the cut-off scores. According to other validation studies which administered other early screening measures on general population (Levante et al., b; Reznick et al., 2007; Ben-Sasson and Carter, 2012), a Q-CHAT total score equal or above the 95th percentile could be applied to detect children at risk for ASD. Based on this criterion, the cut-off score on our sample is 43 and we detected the 29 (5.32%) children at risk in our unselected sample.

5 Discussion

The main purpose of this study was to demonstrate the psychometric and validity properties of the Q-CHAT in an Italian sample of unselected population. Specifically, for the first time, we tested both the Q-CHAT structural validity carrying out a CFA based on the exploratory factor analysis found by Ruta et al. (2019) and we also tested its measurement invariance across gender; we tested the internal validity of the Q-CHAT; finally, we identified the cut-off score to detect children at risk for ASD since no previous studies did it.

The CFA analyses confirmed the 3-factor structure of the Q-CHAT as hypothesized and tested through exploratory factor analysis by Ruta et al. (2019). The three factors extracted are the Non-social/Behavioral autistic traits, the Speech/Language, and Joint attention/non-verbal communication. These aspects represent the three core deficits of ASD as suggested by the DSM 5 (APA, 2013) and several studies (Luyster et al., 2007; Watson et al., 2013; Becerra-Culqui et al., 2018). Specifically, the Non-social/Behavioral autistic traits factor corresponds to the second DSM 5 diagnostic criterion ("*Restricted, repetitive patterns of behavior, interests, or activities*") which manifests itself by repetitive and restricted behaviors and interests, hypo- and hypersensitivity to noise, mannerisms, and difficulties to maintain focus on activities (Matson et al., 2009; Kim and Lord, 2010). The Speech/Language and the Joint attention/non-verbal communication factors refer to the first DSM 5 diagnostic criterion ("*Persistent deficits in social communication and social interaction*") which manifests itself as lack of verbal (see Adamson et al., 2019; Becerra-Culqui et al., 2018) and non-verbal communication (i.e., lack of gestures, and specifically of proto-imperative and declarative pointing; lack of joint attention skill; Watson et al., 2013; Baron-Cohen et al., 1996; Dawson et al., 2004; Adamson et al., 2019). The items 12 and 15 were omitted from the final model tested with the CFA because they had factor loadings lower than .30. The item 12 ("*Does your child place your hand on an object when s/he wants you to use it - e.g. on a door handle when s/he wants you to open the door, on a toy when s/he wants you to activate it*"?), in Ruta e colleagues (2019) EFA, loaded for two different factors (the first, with a plus sign, and the third, with a minus sign). The authors decided to keep the items in their analysis, and this choice, in our opinion, could induce a methodological criticism. For this reason, we decided to delete the item from the analysis.

In the final CFA, three items present some problems. Item 12 had a minus sign and loaded for two factors with factor loadings lower than .30, item 15 had a factor loading lower than .30, and item 18 loaded a factor with a minus sign. The statistical ambiguity of the item 12 could depend on the degree of complexity of this item that could lead to misleading interpretations by parents. A risk marker for ASD is when a child persistently use his/her parent's hand as a tool without sharing gazes with the parent (Franchini et al., 2019). The current formulation of the question does not allow the parents to focus their attention on the presence/absence of the combination of both gaze and gesture. The importance of this issue is also explained and highlighted in the scoring procedure of the Module Toddler and Module 1 of the Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2008), where the integration of gazes and gestures is the marker of a typical developing, whereas atypically developing evaluations are associated to a lack of integration between gazes and gestures. Therefore, our suggestion is to rewording the sentence introducing a reference to the sharing gazes (e.g., "*Does your child place your hand on an object when s/he wants you to use it while look at you in the eyes - e.g. on a door handle when s/he wants you to open the door, on a toy when s/he wants you to activate it*"?).

Regarding the item 15 ("*If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them - e.g. stroking their hair, hugging them*?"), we hypothesize two possible explanations for its exclusion from the analyses. First, the examples reported in the item could be not culturally adequate: for example, in Italy "stroking hair" is not a sign of comfort and this leads parents to misunderstand the meaning of the item. Second, Ruta and colleagues (2019) included this item in the speech and language factor even if its meaning, being empathetic and giving comfort, does not pertain to the language abilities, but to the social communication competence. Thus, we suggest that item 15 would be reworded to meet the cultural specificity of the Italian social context and then tested again as included in the social communicative traits factor.

Finally, the item 18 presented a negative sign on its factor loading for the speech and language factor. According to the modification indices, we forced the item 18 on the Non-social/Behavioral autistic traits factor, but the fit-indices worsened. We have to say that even in the exploratory structure yielded by Ruta and colleagues (2019), the item 18 negatively loaded for the Speech/Language factor. In this vein, our CFA demonstrated that the item 18 has to be reformulated because even if it is worded and measured in the desired direction, that is high scores correspond to high risk of ASD, the sign of the factor loading suggested that the parents interpreted the item in the opposite direction.

Before to test the measurement invariance, we examined the gender difference on Q-CHAT total and factors score. We found significant difference between boys and girls on their total score and on the score of the speech and language factor confirming the fact that there is a difference in the prevalence of ASD by gender (ratio 4:1; APA, 2013). Those results are in line with previous findings on other screening measures of ASD (Levante et al., b) and with evidence about gender differences in autism traits in unselected populations (Constantino and Todd, 2003; Leekam et al., 2007; Reinhardt et al., 2015; Ruzich et al., 2015; Baron-Cohen et al., 2001; Messinger et al., 2015). As regards to the gender invariance, the multi-group CFAs were carried out to determine whether measurement invariance was acceptable for boys and girls and then whether future research in general population could be carried out considering the same factor structure for both of them. The analyses showed the fit measures of the overall model and of the model splitted by gender. As said before, for the overall model not all fit measures are good (CFI and TLI are under the degree of acceptability, while RMSEA, SRMR, GFI and AGFI are acceptable). Our model has a χ^2 significant, low CFI and TLI, and good RMSEA, SRMR, GFI and AGFI. The overall model is used to test configural invariance and we evaluated two further indexes, GFI and AGFI, as Kline (2015) suggested in case of different techniques, as application with complex phenomena with mediation (Fabio et al., 2005, 2018; Gunzler et al., 2013; Ingusci et al., 2016), moderation and latent variables. As reported in Putnick and Bornstein (2016), configural invariance is tested by evaluating the overall fit of the model. In this study, we can figure that the model has a tolerable configural invariance. By considering the models divided by boys and girls, we found that the fit was slightly better for girls than boys with, again, low CFI and TLI, and good RMSEA and SRMR. In testing invariance in our model, we found inconsistence between CFI and RMSEA indices. As suggested by Lai and Green (2016) this situation can depend on three causes: the misunderstanding of the meaning of "good fit" and "bad fit" in the current literature, the dissimilar point of views about the cutoff values and the different perspectives under the two indices. We decided to focus on this latter motivation. More specifically, in our case RMSEA has an acceptable value; on the contrary CFI does not respect this condition. RMSEA index aim is to evaluate the computed error when CFA try to approximate the estimated parameters with the observed values. It is a fit measure that test how far a hypothesized model is from a perfect model. Instead CFI is an incremental index that compare the fit of a hypothesized model with a baseline one, with no relations between variables. In our model we found that Item 12 has a similar loading with two factors (1 and 3) and this can be read as a significant cross-loading. CFI can be influenced by inter-correlations, because the fit of the baseline model improves and consequently the one for the hypothesized model decreases. The same argumentation can be proposed referring to reliability, as Factor 2 has a low McDonald's Omega coefficient. One of the reasons of this low reliability index could be the presenze of further inter-correlations. One future direction could be to deepen investigate the cross-loadings between variables and dimensions.

As mentioned in Table 2, looking at the differences of p. values for the Δchi-squares, we can assume that there is non-invariance of factor loadings and residuals. However, since the chi-square and Δchi-square are indices influenced by the size of the sample and the complexity of the model (it become less accurate when the sample size increases), we preferred the ΔCFI and ΔRMSEA to test the invariance. According to these indexes, we found all models values for ΔRMSEA smaller than .015 and equal or smaller than .010 for the ΔCFI, except for the residual model. We can assume that the model had non-invariant residuals, which are the sum of variance not shared with the factor and measurement error. Residuals non-invariance implies that at least one item residual is different across groups.

Our third aim was to examine the internal validity of the screening measure. The correlations found among the three factors of the Q-CHAT are demonstration of internal validity of the questionnaire.

Finally, we calculated the cut-off score to detect children at risk. As reported in the results section, we considered the score at or above the 95th percentile as the risk cut-off. Since no previous validation studies examine this statistical parameter, our finding represents the first contribution in this sense. Future studies are needed to test the reliability of this cut-off score to detect children with ASD.

The study has two main limitations. First, the sample was recruited in one region located in South Italy. Although that could have led to a sampling bias with implications on the generalization of the results, we have to say that the comparisons between our findings and those obtained by other authors (Ruta et al., 2019), who collected data in other regions of Italy, are similar. Therefore, the bias determined by the location of the sample could be considered as negligible.

The second limitation pertains the fact that this study reports cross-sectional data from a longitudinal study, which is still in progress. In the future, it would be possible to assess the Q-CHAT test-retest reliability, criterion validity, its PPV and NPV. This, will allow to explore the full range of the psychometric properties of the measure. Another study limitation is related to the response rate of the pediatricians (close to 50% of the total). The parents' response rate was quite high (86.1%), meaning that, when the pediatricians' participation was secured, the parental agreement was high. Probably this pattern may be explained by the quality of communication between parents and doctors and their reciprocal trust (Petrocchi et al., 2018, 2019) and suggested that researchers should implement interventions on this topic.

6 Conclusion

In conclusion, the present study contributes to the extension of the literature on the early screening of ASD testing, for the first time, the structural validity applying the CFA, whereas previous studies are limited because explored the structure of the Q-CHAT via exploratory factor analysis. The present paper gave also demonstration of gender invariance which is an important feature of the Q-CHAT given the male: female ratio of the ASD. From a clinical point of view, the present research suggests a cut-off score for the application of the Q-CHAT for prevention and early identification of ASD. This set of positive features of the Q-CHAT allow to rec- ommend it for the application during the pediatric surveillance in order to detect the early signs of risk of ASD in children from 18 to 24 months of life.

The modest demonstration of internal consistency needs further explorations, as well as the evaluation of the criterion validity of the Q-CHAT comparing its scores with those yielded by the so-called gold-standard measures for the ASD diagnosis (e.g., the Autism Diagnostic Interview-Revised, Lord et al., 1994; the Autism Diagnostic Observation Schedule-2, Lord et al., 2008). The early detection of atypical developmental trajectories may support medical decision-making on further steps for child referral to an early diagnostic assessment which may enable early interventions when needed (Anderson et al., 2014; Daniels and Mandell, 2014; Dawson et al., 2010; Perry et al., 2008; Renty and Roeyers, 2006; Robins and Dumont-Mathieu, 2006; Sallows and Graupner, 2005; Sutera et al., 2007). This early screening, diagnosis, and intervention could improve better children and family outcomes (Lecciso et al., 2013).

Conflict of interests: none. **Funding:** none. **Author's contribution:** FL and SP conceived the research project and AL participated in the design definition. FL, SP, and AL wrote the paper. AL and FS performed the statistical analyses and wrote the results section. FL coordinated the research team. All the authors read and approved the final version of the manuscript.

Acknowledgement

We are very grateful to the local public health service and the two Associations of Pediatricians, the Italian Federation of Medical Pediatricians (FIMP) and the Italian Confederation of Pediatrics (CIPe) for their agreement to the study. Specifically, we are grateful to dr. Serafino De Giorgi, dr. Angelo Massagli, and dr. Maria Rosaria Filograna for their essential collaboration during all the phases of the research. We are also very grateful to the pediatricians and parents for their participation and collaboration during the data collection. Finally, we are grateful to the psychologists Filomena De Lumè, Luigia Duma, Valentina Garrapa, Emilia Perrone, Viviana Vetrugno, Annamaria Vizzi, Giuseppe Antonioli for their collaboration in data collection.

References

Achenbach, T. M. and Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles*, volume 30. Burlington, VT: University of Vermont, Research center for children, youth and families.

Adamson, L. B., Bakeman, R., Suma, K., and Robins, D. L. (2019). An expanded view of joint attention: Skill, engagement, and language in typical development and autism. *Child development*, 90(1):e1–e18.

Allison, C., Baron-Cohen, S., Wheelwright, S., Charman, T., Richler, J., Pasco, G., and Brayne, C. (2008). The q-chat (quantitative checklist for autism in toddlers): A normally distributed quantitative measure of autistic traits at 18–24 months of age: Preliminary report. *Journal of autism and developmental disorders*, 38(8):1414–1425.

Anderson, D. K., Liang, J. W., and Lord, C. (2014). Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, 55(5):485–494.

APA (2013). Diagnostic and statistical manual of mental disorders. *BMC Med*, 17:133–137.

APA, A. P. A. (1994). Diagnostic and statistical manual of mental disorders. *DSM-IV*.

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

Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N., Morgan, K., Drew, A., and Charman, T. (1996). Psychological markers in the detection of autism in infancy in a large population. *The British Journal of Psychiatry*, 168(2):158–163.

Baron-Cohen, S., Wheelwright, S., Cox, A., Baird, G., Charman, T., Swettenham, J., Drew, A., and Doehring, P. (2000). Early identification of autism by the checklist for autism in toddlers (chat). *Journal of the royal society of medicine*, 93(10):521–525.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (aq): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders*, 31(1):5–17.

Becerra-Culqui, T. A., Lynch, F. L., Owen-Smith, A. A., Spitzer, J., and Croen, L. A. (2018). Parental first concerns and timing of autism spectrum disorder diagnosis. *Journal of autism and developmental disorders*, 48(10):3367–3376.

Ben-Sasson, A. and Carter, A. S. (2012). The application of the first year inventory for asd screening in israel. *Journal of autism and developmental disorders*, 42(9):1906–1916.

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, 16(3):201–213.

Byrne, B. (2010). Structural equation modeling with amos: Basic concepts, applications and programming.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., and Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in rmsea test statistic in structural equation models. *Sociological methods & research*, 36(4):462–494.

Cheung, G. W. and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2):233–255.

Ciavolino, E. (2012). General distress as second order latent variable estimated through pls-pm approach. *Electronic Journal of Applied Statistical Analysis*, 5(3):458–464.

Constantino, J. N. and Todd, R. D. (2003). Autistic traits in the general population: a twin study. *Archives of general psychiatry*, 60(5):524–530.

Crisci, A. and D'Ambra, A. (2012). Permutation test for group comparison in pls-path modeling. *Electronic Journal of Applied Statistical Analysis*, 5(3):339–345.

Daniels, A. M. and Mandell, D. S. (2014). Explaining differences in age at autism spectrum disorder diagnosis: A critical review. *Autism*, 18(5):583–597.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., Donaldson, A., and Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model. *Pediatrics*, 125(1):e17–e23.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., and Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental psychology*, 40(2):271.

Fabio, R., Martinazzoli, C., and Antonietti, A. (2005). Costruzione e standardizzazione dello strumento "rars"(rett assessment rating scale). *Ciclo Evol Disabil*, 8:257–281.

Fabio, R. A., Caprì, T., Mohammadhasani, N., Gangemi, A., Gagliano, A., and Martino, G. (2018). Frequency bands in seeing and remembering: comparing ADHD and typically developing children. *Neuropsychological Trends*, 24:97–116,
DOI: *10.7358/neur-2018-024-fabio*.

Fabio, R. A., Oliva, P., and Murdaca, A. M. (2011). Systematic and emotional contents in overselectivity processes in autism. *Research in Autism Spectrum Disorders*, 5(1):575–583.

Franchini, M., Armstrong, V. L., Schaer, M., and Smith, I. M. (2019). Initiation of joint attention and related visual attention processes in infants with autism spectrum disorder: Literature review. *Child Neuropsychology*, 25(3):287–317.

Gunzler, D., Chen, T., Wu, P., and Zhang, H. (2013). Introduction to mediation analysis with structural equation modeling. *Shanghai archives of psychiatry*, 25(6):390.

Ingusci, E., Manuti, A., and Callea, A. (2016). Employability as mediator in the relationship between the meaning of working and job search behaviours during unemployment.

Electronic Journal of Applied Statistical Analysis, 9(1):1–16.

Kelloway, E. K. (1995). Structural equation modelling in perspective. *Journal of Organizational Behavior*, 16(3):215–224.

Kim, S. H. and Lord, C. (2010). Restricted and repetitive behaviors in toddlers and preschoolers with autism spectrum disorders based on the autism diagnostic observation schedule (ados). *Autism Research*, 3(4):162–173.

Kline, R. B. (2015). *Principles and practice of structural equation modeling.* Guilford publications.

Lai, K. and Green, S. B. (2016). The problem with having two watches: Assessment of fit when rmsea and cfi disagree. *Multivariate behavioral research*, 51(2-3):220–239.

Lecciso, F., Petrocchi, S., Savazzi, F., Marchetti, A., Nobile, M., and Molteni, M. (2013). The association between maternal resolution of the diagnosis of autism, maternal mental representations of the relationship with the child, and children's attachment. *Life Span and Disability*, 16(1):21–38.

Leekam, S., Tandos, J., McConachie, H., Meins, E., Parkinson, K., Wright, C., Turner, M., Arnott, B., Vittorini, L., and Couteur, A. L. (2007). Repetitive behaviours in typically developing 2-year-olds. *Journal of Child Psychology and Psychiatry*, 48(11):1131–1138.

Levante, A., Petrocchi, and Lecciso, F. (2017). Quantitative-checklist for autism in toddler (q-chat) italian translation. *Retrieved from https : //www.autismresearchcentre.com/arc_tests.*.

Levante, A., Petrocchi, S., and Lecciso, F. Systematic review of measures for the early detection of risk of autism spectrum disorder risk in toddlers. Under review.

Levante, A., Petrocchi, S., and Lecciso, F. (2019). Systematic review protocol of measures for the early detection of risk of autism spectrum disorder risk in toddlers. *Lifespan and Disabilities*, 22(1).

Levante, A., Petrocchi, S., Massagli, A., M.R., F., and Lecciso, F. The early screening of the autism spectrum disorder: Validity properties and cross-cultural generalisability of the first year inventory in italy. Under review.

Lord, C., Rutter, M., Dilavore, P. C., and Risi, S. (2008). *ADOS: Autism diagnostic observation schedule.* Hogrefe Boston, MA.

Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders*, 24(5):659–685.

Luyster, R., Lopez, K., and Lord, C. (2007). Characterizing communicative development in children referred for autism spectrum disorders using the macarthur-bates communicative development inventory (cdi). *Journal of Child Language*, 34(3):623–654.

Magiati, I., Goh, D. A., Lim, S. J., Gan, D. Z. Q., Leong, J., Allison, C., Baron-Cohen, S., Rifkin-Graboi, A., Broekman, B. P., Saw, S. M., et al. (2015). The psychometric properties of the quantitative-checklist for autism in toddlers (q-chat) as a measure of autistic traits in a community sample of singaporean infants and toddlers. *Molecular autism*, 6(1):40.

Matson, J. L., Dempsey, T., and Fodstad, J. C. (2009). Stereotypies and repetitive/restrictive behaviours in infants with autism and pervasive developmental disorder. *Developmental Neurorehabilitation*, 12(3):122–127.

Messinger, D. S., Young, G. S., Webb, S. J., Ozonoff, S., Bryson, S. E., Carter, A., Carver, L., Charman, T., Chawarska, K., Curtin, S., et al. (2015). Early sex differences are not autism-specific: A baby siblings research consortium (bsrc) study. *Molecular autism*, 6(1):32.

Perry, A., Cummings, A., Geier, J. D., Freeman, N. L., Hughes, S., LaRose, L., Managhan, T., Reitzel, J.-A., and Williams, J. (2008). Effectiveness of intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders*, 2(4):621–642.

Petrocchi, S., Iannello, P., Lecciso, F., Levante, A., Antonietti, A., and Schulz, P. (2019). Interpersonal trust in doctor-patient relation: Evidence from dyadic analysis and association with quality of dyadic communication. *Social Science & Medicine*, 235:112391.

Petrocchi, S., Labrie, N. H., and Schulz, P. J. (2017). Measurement invariance of the short wake forest physician trust scale and of the health empowerment scale in german and french women. *Journal of health psychology*, page 1359105317719582.

Petrocchi, S., Rotenberg, K. J., Levante, A., and Lecciso, F. (2018). Children's trust in social workers: Scale development and relations to children's engagement with social workers. *Child & Family Social Work*, 23(2):239–247.

Pornprasertmanit, S., Miller, P., Schoemann, A., Rosseel, Y., Quick, C., Garnier-Villarreal, M., Selig, J., Boulton, A., Preacher, K., Coffman, D., et al. (2015). Package 'semtools'. *Retrieved in.*

Putnick, D. L. and Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental review*, 41:71–90.

Reinhardt, V. P., Wetherby, A. M., Schatschneider, C., and Lord, C. (2015). Examination of sex differences in a large sample of young children with autism spectrum disorder and typical development. *Journal of autism and developmental disorders*, 45(3):697–706.

Renty, J. and Roeyers, H. (2006). Quality of life in high-functioning adults with autism spectrum disorder: The predictive value of disability and support characteristics. *Autism*, 10(5):511–524.

Reznick, J. S., Baranek, G. T., Reavis, S., Watson, L. R., and Crais, E. R. (2007). A parent-report instrument for identifying one-year-olds at risk for an eventual diagnosis of autism: the first year inventory. *Journal of autism and developmental disorders*, 37(9):1691–1710.

Robins, D. L. and Dumont-Mathieu, T. M. (2006). Early screening for autism spectrum disorders: update on the modified checklist for autism in toddlers and other measures. *Journal of Developmental & Behavioral Pediatrics*, 27(2):S111–S119.

Robins, D. L., Fein, D., Barton, M. L., and Green, J. A. (2001). The modified checklist for autism in toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of autism and developmental disorders*, 31(2):131–144.

Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling and more. version 0.5–12 (beta). *Journal of statistical software*, 48(2):1–36.

Ruta, L., Arduino, G. M., Gagliano, A., Apicella, F., Leonardi, E., Famà, F. I., Chericoni, N., Costanzo, V., Turco, N., Tartarisco, G., et al. (2019). Psychometric properties, factor structure and cross-cultural validity of the quantitative checklist for autism in toddlers (q-chat) in an italian community setting. *Research in Autism Spectrum Disorders*, 64:39–48.

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., and Baron-Cohen, S. (2015). Measuring autistic traits in the general population: a systematic review of the autism-spectrum quotient (aq) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular autism*, 6(1):2.

Sallows, G. O. and Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American journal on mental retardation*, 110(6):417–438.

Schafer, J. L. and Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2):147.

Sutera, S., Pandey, J., Esser, E. L., Rosenthal, M. A., Wilson, L. B., Barton, M., Green, J., Hodgson, S., Robins, D. L., Dumont-Mathieu, T., et al. (2007). Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of autism and developmental disorders*, 37(1):98–107.

Watson, L. R., Crais, E. R., Baranek, G. T., Dykstra, J. R., and Wilson, K. P. (2013). Communicative gesture use in infants with and without autism: A retrospective home video study. *American Journal of Speech-Language Pathology*.

WHO (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*, volume 2. World Health Organization.

Zwaigenbaum, L., Bauman, M. L., Choueiri, R., Kasari, C., Carter, A., Granpeesheh, D., Mailloux, Z., Roley, S. S., Wagner, S., Fein, D., et al. (2015). Early intervention for children with autism spectrum disorder under 3 years of age: recommendations for practice and research. *Pediatrics*, 136(Supplement 1):S60–S81.