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/i20705948v14n1p166

Test-retest reliability and measurement invariance across time of the Quantitative-CHecklist for autism in toddler

By Levante, Petrocchi, Lecciso

Published: 20 May 2021

Electronic Journal of Applied Statistical Analysis
Vol. 14, Issue 01, May 2021, 146-166
DOI: 10.1285/i20705948v14n1p166

Test-retest reliability and measurement invariance across time of the Quantitative-CHecklist for autism in toddler

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

[a] *Department of History, Society, and Human Studies, University of Salento, Via di Valesio, 73100 Lecce, Italy*
[b] *Lab of Applied Psychology, Department of History, Society, and Human Studies, University of Salento, 73100 Lecce, Italy*
[c] *USI, Institute of Communication and Health, Via Buffi 6, 6900 Lugano, Switzerland*

Published: 20 May 2021

The study evaluated the stability over time of the Q-CHAT (i.e., Quantitative-CHecklist for Autism in Toddler), a screening tool for the early detection of Autism. The Q-CHAT is filled by parents when the child is from 18-36 months old, and whether a risk score is found, the clinician refers the child to a more in-depth diagnostic evaluation. Previous studies demonstrated the measure test-retest reliability on a small time interval (1-6 months). Therefore, the present study tested the stability of the scores considering a larger time interval of 18 months. No previous studies have tested its measurement invariance over time. Since the tool is filled by parents during a large time span (18-36 months of life), it is imperative to understand whether there is an invariance over time in their evaluations. The Italian version of the Q-CHAT was completed by 282 parents of children with no pre-existing signs of risk of Autism. The Q-CHAT was administered when children were 18 months (T1) and then 18 months later (T2). The intraclass correlation coefficients for the test-retest reliability ranged from sufficient to moderate. The measurement invariance across time revealed a tolerable configural and metric invariance. Contrary, the scalar invariance was not met meaning that the means of the constructs are not invariant over time. The findings give a further demonstration of the reliability of the Q-CHAT. They give evidence that changes

*Corresponding author: annalisa.levante@unisalento.it

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

in the scores would reflect real changes in the construct itself, and not in the way individuals interpret the measure items.

keywords: autism spectrum disorder, Quantitative-CHecklist for Autism in Toddler, Q-CHAT, test-retest reliability, measurement invariance across time, validation.

1. Introduction

In the last two decades, several screening tools have been developed to detect early warning signs of risk for Autism Spectrum Disorder (ASD), which is a neurodevelopmental disorder (APA, 2013). Those measures are mostly parent-reported questionnaires, administered from the first months of life and demonstrating adequate validity and psychometric sound characteristics according to recent systematic reviews (Petrocchi et al., 2020b; Thabtah and Peebles, 2019). Among them, the Quantitative-CHecklist for Autism in Toddler (Allison et al., 2008) was developed to quantify the autistic traits in children aged 18-36 months, shifting the binary scoring items (yes/no) of the first versions of the questionnaire (Robins et al., 2014, 2001; Baron-Cohen et al., 1992) in the final 4-point Likert scale. As the Q-CHAT developer argued (Allison et al., 2008), the dichotomous scoring is more stringent and conservative, whereas the frequency evaluation of several key behaviors (e.g., pointing, pretend play, communication and language, repetitive behaviors) allows clinicians to detect mild cases of ASD. Furthermore, the risk threshold was also revised. In the previous versions of the questionnaire (Robins et al., 2014, 2001; Baron-Cohen et al., 1992), the cut-off was evaluated considering whether three specific key behaviors (i.e., pretend play, pointing, and joint attention) were present or absent. Whereas the latest version of the questionnaire, i.e. the Q-CHAT (Allison et al., 2008), evaluates the scores along a continuum of ASD symptoms severity calculating three different scores (or factors) and a final total score.

Several Q-CHAT psychometric properties have been tested in the general population (Lecciso et al., 2019; Ruta et al., 2019a; Magiati et al., 2015; Allison et al., 2008) and clinical groups (Devescovi et al., 2020; Ruta et al., 2019b). However, further studies are needed to extend the knowledge about the stability of the scores over time. In this vein, the present study purpose was to test two specific psychometric properties of the Q-CHAT, that is the test-retest reliability and the measurement invariance across time. Both properties focus the attention on the stability of the Q-CHAT over time and they are pivotal to measure the global reliability of the tool.

To investigate the Q-CHAT test-retest reliability, four studies have been conducted (Park et al., 2018; Magiati et al., 2015; Mohammadian et al., 2015; Allison et al., 2008). The first validation study (Allison et al., 2008) examined the psychometric property on English general population ($n = 330$). The time interval between the two administrations was 38 days ($sd = 12$, range 15–109) and the intraclass correlation coefficient revealed a good reliability ($r = .82$ for single measures). The study by Magiati et al. (2015) examined the test-retest reliability on Singaporean community samples with a time interval

between the two administrations of 6 months. The questionnaires were completed by parents of two groups of children aged 18 ($n = 368$) and 24 ($n = 396$) months and found a moderate test-retest reliability ($r = .60$ and $r = .64$ respectively). The Persian translation of the tool was conducted by Mohammadian et al. (2015). They calculated the test-retest reliability of the Q-CHAT on a small sample ($n = 30$ children with ASD) with a time interval of 1 month funding an excellent value ($r = .99$). Finally, the recent validation study by Park et al. (2018) evaluated the Korean version of the screening tool. The time interval between the two administrations was 4 weeks and the test-retest reliability was calculated on a small sample ($n = 20$) with a good value ($r = .83$).

Those previous validation studies of the Q-CHAT applied a time interval ranging from 1 to 6 months, whereas no studies examined the test-retest reliability applying a larger time interval. Testing the stability of its scores over a larger time interval contributes to the increasing of the knowledge of the reliability of the measurement and the methodological rigor applied to operationalize the construct itself (Heal and Twycross, 2015). Furthermore, the Q-CHAT scores are used to detect early signs of risk of Autism and to refer the child to a more in depth diagnostic evaluation whether a risk is found. Demonstration of the stability of the scores over a long time interval is crucial to be confident that they actually reflect the child's skills and behaviors.

Therefore, the first aim of the present research was to evaluate the Q-CHAT test-retest reliability considering a larger time interval compared to the ones applied in the previous validation studies. Since the Q-CHAT measures the frequency of several key behaviors, which are subjected to change during childhood, we expected to find two different magnitudes of the reliability. We expected to reach a moderate stability for factor 1, evaluating the child's non-social/behavioral autistic traits (i.e., repetitive and restricted behaviors), because the frequency of those behaviors decreases only after 3/4 years of age (Istvan et al., 2020; Harrop et al., 2014; Honey et al., 2007; Leekam et al., 2007). Moreover, we expected to find sufficient stability for factor 2, measuring speech and language, and for factor 3, evaluating joint attention and non-verbal communication, because the behaviors included in the evaluation should show a developmental change in the time interval considered (i.e, 18-36 months) (Benítez-Burraco and Progovac, 2020; Lameira and Call, 2020; Capirci et al., 2005; Volterra, 1981).

The previous mentioned studies on the Q-CHAT (Park et al., 2018; Magiati et al., 2015; Mohammadian et al., 2015; Allison et al., 2008) did not test its measurement invariance, with the exception of the gender invariance (Lecciso et al., 2019). Therefore, studies in this field are particularly needed. The measurement invariance requires the assessment of whether the measurement parameters of latent variables remain invariant (Chang et al., 2014; Widaman et al., 2010) across gender, culture, or time, as in our case, regarding to their factor loadings (i.e., threshold invariance), intercepts (i.e., metric invariance), and construct (i.e., scalar invariance). In particular, the measurement invariance across time allows researchers being confident that the hypotheses and the inferences about changes in the measure scores would reflect real changes in the construct itself (Olino, 2020; Tyrell et al., 2019), and not in the way individuals interpret the items of the questionnaire. Therefore, measurement invariance and longitudinal analysis (Marzorati et al., 2020, 2019) is a pivotal psychometric property to improve the robustness of mea-

sures (Lee, 2018) especially when evaluating individuals' psychopathology (Molino et al., 2020) and mental health (Petrocchi et al., 2020a).
The present research intended to test measurement invariance of the Q-CHAT over time. In other words, our second aim was to demonstrate whether the way in which parents interpret the meaning of the Q-CHAT items remain invariant over time to be confident that what eventually is changing is the real scores, so the real behaviors and competencies shown by their children. Summarizing, the present paper aimed to test the following hypotheses and research question:
Hypothesis 1. *It was expected that the evaluation of the child's non-social/behavioral autistic traits (Factor 1) would show a moderate stability over time.*
Hypothesis 2. *It was expected that the evaluation of the child's speech and language (Factor 2) and joint attention/non-verbal communication (Factor 3) skills would show a sufficient stability over time.*
Research Question 1. *Are the Q-CHAT scores stable over time?.*

2. Materials and Method

2.1. Procedure

The study reported data from a bigger longitudinal project. One hundred fifty-five pediatricians working for the Local Public Health Service in a big city in the South of Italy were contacted and sixty-four of them (55.6%) agreed to participate in the study. All families with a child born from February to September 2016 were invited and participants signed a consent form. Exclusion criteria comprised any children's pre-existing medical conditions. In the present paper, we considered the data collected at a two-time point, from September 2017 to May 2018 (T1) and from January to October 2019 (T2). At T1 the Q-CHAT was administered for the first time to parents of children aged 18 months, and at T2 the same parents filled the Q-CHAT for the second time. The Ethical Committee of the Local Health Service gave its approval (n. 528/8).

2.2. Participants

The sample included 282 questionnaires completed by parents of children aged 18 (T1) and when they were 36 months (T2). One hundred forty-nine (52.8%) of the questionnaires were completed for males and 131 (46.5%) for females. The children's mean age was 18.3 months ($sd = 1$); 204 (72.3%) children were born at term (after the 38th gestation weeks) and 65 (23%) were born preterm. One hundred eight (38.3%) were first-born and 140 (49.7%) were second-born or more. The questionnaires were completed by mother for 71.6% ($n = 202$), by fathers for 5.3% ($n = 15$), and for 23% ($n = 65$) by both parents. Mother's mean age was 34.3 years ($sd = 5.6$ years) and the educational level was low (up to 8 years) for 24.8% ($n = 70$), intermediate (up to 13 years) for 41.1% ($n = 116$), and high for 30.5% ($n = 86$). Father's mean age was 37.9 years ($sd = 6.6$ years) and the educational level was low for 35.8% ($n = 101$), intermediate for 39.7% ($n = 112$), and high for 18.1% ($n = 51$).

2.3. Measure

The Italian version of Q-CHAT (see Appendix) was translated by Levante, Petrocchi, and Lecciso (2017) and approved by the original authors (Allison et al., 2008). The original version of the Q-CHAT includes 25 items and half of them are reverse scored (items 3, 7, 8, 11-13, 16, 18, 20, 22-25). The Q-CHAT evaluates the frequency of several crucial behaviors of the children' functioning, i.e., joint attention, social and pretend play, social interest, imperative and declarative pointing, language development, repetitive behaviors, and non-verbal behaviors. Response options are rated from 0 ("always") to 4 ("never"). Three partial scores assessing the autistic traits have been created according to the confirmatory factor structure found by Lecciso et al. (2019): i) Non-social/Behavioral autistic traits (theoretical range: 0-36; T1: $\alpha = .79$, ITC $\geq .21$; T2: $\alpha = .79$, ITC $\geq .22$); ii) Speech and Language (theoretical range: 0-20; T1: $\alpha = .86$, ITC $\geq .19$; T2: $\alpha = .22$, ITC $\geq .11$); and iii) Joint attention/Non-verbal communication (theoretical range: 0-16; T1: $\alpha = .40$, ITC $\geq .14$; T2: $\alpha = .69$, ITC $\geq .21$). Finally, a total score was created as a sum of items (theoretical range: 0-100; T1: $\alpha = .69$, ITC $\geq .25$; T2: $\alpha = .57$, ITC $\geq .22$). Higher scores indicate greater risk of autism.

2.4. Statistical Analysis

The statistical analyses were performed using Jamovi (Version 1.2) and RStudio (R Core Team, 2013). The Cronbach' alpha and the Item-Total Correlation (ITC) were calculated for scores collected on T0 and T1. The Q-CHAT scores distribution was checked. The IntraClass Correlation analysis was conducted for the three factors and the total score. According to Koo and Li (2016) the test-retest reliability can be interpreted as a moderate if the ICC value is between to .50 and .75, good when the ICC is between .75 to .90, and excellent if the ICC value is above .90. To investigate the measurement invariance across time the Lavaan (Rosseel, 2012) and SemTools (Pornprasertmanit et al., 2015) packages were applied. Four nested models were compared:

a. Model 1: configural model;

b. Model 2: model with factor loadings constrained as equal across time (threshold model);

c. Model 3: model with factor loadings and item intercepts constrained as equal across time (metric model);

d. Model 4: model with factor loadings, item intercepts, and construct means constrained as equal across time (scalar model).

As recommended by Hu and Bentler (1999) and Cheung and Rensvold (2000), a change in CFI of less than .01 is evidence of invariance.

3. Results

3.1. Preliminary analyses

Preliminary analyses showed that the variables had less than 5% of missing data. As recommended by the original authors (Allison et al., 2008) and as done by others (Ruta et al., 2019a; Magiati et al., 2015), missing data were imputed with 0. At T1, that is when children were 18-21 months old, the Q-CHAT total score (M = 28.5; sd = 8.5; range = 9-56) was normally distributed with skewness = .163 and kurtosis = .139. The Non-social/behavioral autistic traits factor mean was 12.4 (sd = 6.6; range = 0-33); the Speech and Language factor mean score was 7.1 (sd = 1.8; range = 1-14); finally, the Joint attention/Non-verbal communication factor mean score was .92 (sd = 1.3; range = 0-8). At T2, when children were 34-36 months old, the Q-CHAT total score (M = 22.8; sd = 8.1; range = 4-48) was normally distributed with skewness = .388 and kurtosis = .281. The Non-social/behavioral autistic traits mean was 9.2 (sd = 5.9; range = 0-31); the Speech and Language factors mean was 5.4 (sd = 1.7; range = 0-13); finally, the Joint attention/Non-verbal communication factor mean score was 1.8 (sd = 2.3; range = 0-13).

3.2. Test-retest Reliability

The time interval between the two Q-CHAT administrations ranged 18 months and the intraclass correlation coefficients (ICC) were calculated for the total score and for each factor. The ICC between the two total scores was α = .63 for single measure (lower CI = .36; upper CI = .55). The ICC for single measure between the factor-scores were: α = .63 (lower CI = .36; upper CI = .55) for Factor 1, α = .42 (lower CI = .16; upper CI = .38) for Factor 2, and α = .40 (lower CI = .14; upper CI = .36) for Factor 3.

3.3. Measurement Invariance

The three-factor structure confirmed by Lecciso et al. (2019) was tested here for measurement invariance across time. As done by others (Chang et al., 2014), we tested the measurement invariance across time for the three factors separately to avoid violation of the principle of parsimony of the CFA. Results showed a tolerable configural invariance for each factor meaning that it is appropriate to go on and test other levels of invariance (Table1). The results are quite similar for all the three factors. The invariance of the threshold models was met for all the three factors, as well as for the metric models. As for the scalar invariance, both the $\Delta\chi^2$ and the ΔCFI show that there is no invariance. Therefore, the means of the constructs are not invariant over time.

[H]

Table 1: Data fit indices and model comparison for the three factor of the Q-CHAT

Model	χ^2	**DF**	**CFI**	**RMSEA**	**Comparison**	$\Delta\chi^2$	ΔDF	ΔCFI	$\Delta RMSEA$
Factor 1									
Configural model	221.155***	125	.92	.05	-	-	-	-	-
Threshold model	221.155***	125	.92	.05	Threshold-Configural	0	0	0	0
Metric model	221.155***	125	.92	.05	Metric-Threshold	0	0	0	0
Scalar model	347.556**	133	.82	.08	Scalar-Metric	126.4**	8	-.1	.025
Factor 2									
Configural model	59.699***	29	.91	.06	-	-	-	-	-
Threshold model	59.699***	29	.91	.06	Threshold-Configural	0	0	0	0
Metric model	59.699***	29	.91	.06	Metric-Threshold	0	0	0	0
Scalar model	87.45**	33	.86	.08	Scalar-Metric	27.445**	4	-.06	.02
Factor 3									
Configural model	28.497*	15	.94	.06	-	-	-	-	-
Threshold model	28.497*	15	.94	.06	Threshold-Configural	0	0	0	0
Metric model	28.497*	15	.94	.06	Metric-Threshold	0	0	0	0
Scalar model	43.357**	28	.89	.07	Scalar-Metric	14.86*	3	-.05	.01

* $p \leq .05$; * $p \leq .01$; ** $p \leq .001$.

Note: Df: degree of freedom; CFI: comparative fit index; RMSEA: root mean square of approximation; SRMR: standardized root mean square residual. Factor 1: Non-social/behavioral autistic traits; Factor 2: Speech and Language; Fator 3: Joint attention/Non-verbal communication

4. Discussion

The main purpose of the present research was to investigate the stability over time of the Q-CHAT. Specifically, we examined the test-retest reliability and the measurement invariance across time in an unselected sample of Italian toddlers. To explore whether and how the autistic traits severity measured by the Q-CHAT changes over time in the general population, we examined the test-retest reliability considering a wider time interval than other previous validation studies (Park et al., 2018; Magiati et al., 2015; Mohammadian et al., 2015; Allison et al., 2008). We formulated two hypotheses, corroborated by our results. We found a moderate reliability coefficient related to the factor 1, measuring non-social/behavioral autistic traits (i.e, repetitive and restrictive behaviors), and a sufficient reliability regarding the other two factors, measuring speech and language, and joint attention/non-verbal communication.

We expected a moderate reliability for factor 1 because the frequency of non-social behaviors (e.g., line objects up; focus attention on spinning objects; repetitive behaviors; restricted interests) tends to decrease only after the age of 3/4 in typically developing children (Istvan et al., 2020; Harrop et al., 2014; Honey et al., 2007; Leekam et al., 2007; Tregay et al., 2009). Therefore, we expected that parents, who identified on the Q-CHAT those behaviors when children were 18 months old, then identified similarly the same behaviors at 36 months. Similarly, we expected that behaviors under the speech and language and joint attention/non-verbal communication categories would be less stable over a time of 18 months because those behaviors show expected developmental changes in this period (Benítez-Burraco and Progovac, 2020; Lameira and Call, 2020; Capirci et al., 2005; Volterra, 1981). Therefore, the results of the present study gave further evidence of the test-retest reliability of the Q-CHAT over a longer period compared to the one applied in the previous research. Our results increase the demonstrations of the reliability of the measurement and the methodological rigor applied to operationalize the construct (Heal and Twycross, 2015).

The second purpose was to test, for the first time, the measurement invariance of the Q-CHAT over time. Our results demonstrated the configural invariance for the three factors meaning that the basic organization of the constructs is supported in the two time-point. We also found a demonstration of the metric invariance for the three factors, indicating that each item contributes to the latent construct to a similar degree across the two time-point. In our results, the overall fit of the metric models for the three factors were significantly better, compared to the configural invariance models. In other words, all the factor loadings are related to the constructs consistently over time. Finally, our results demonstrated that even the metric invariance of the three factors was met, meaning that even the item intercepts are invariant over time. We did not find a demonstration of the scalar invariance of the three factors, indicating that the scores of the three factors change over time.

Overall considered, our findings allow researchers being confident that the hypotheses and the inferences about changes in the measure scores would reflect real changes in the construct itself (Tyrell et al., 2019), and not in the way individuals interpret the items of the questionnaire. The autistic traits, measured through the three factors of

the Q-CHAT, assume the same meaning across the two measurements (Putnick et al., 2016).

One main limitation of the study is the geographical definition of the sample that may have limited the generalizability of the results. Although the study should be replicated with a large sample in a larger geographical area, these results had an important clinical implication in the fact of consistently identifying children at risk over time. These preliminary results contribute to the debate regarding the most appropriate measure to apply for detecting warning signs of risk of autism during pediatric surveillance (Petrocchi et al., 2020b; Levante et al., 2019; Thabtah and Peebles, 2019). Testing and demostrating that screenng tools arestable and reliable is pivotal: they should allow to reveal the mild changes over time of the autistic traits along the child development trajectories, as it has been done for the *gold standard* diagnostic instrument for autism (Lord et al., 2012). Indeed, according to several authors (Anderson et al., 2014; Dawson et al., 2010; Perry et al., 2008; Renty and Roeyers, 2006; Robins and Dumont-Mathieu, 2006), the early detection of risk conditions in the general population allows the clinician to send the child to an early diagnostic procedure and early treatment (Leo et al., 2019, 2018))with a better prognosis (Rodgers et al., 2020; Sinai-Gavrilov et al., 2020; Haine-Schlagel et al., 2020; Lecciso et al., 2013).

Confict of interests: none. **Funding**: none.

Author's contribution: FL conceive the study. FL, SP, and AL designed the study methodology and AL collected data. AL analyzed the data and wrote the draft paper; SP and FL revised the manuscript and gave feedback. All authors read and approved the final version of the manuscript.

Acknowlegdement: We thank the pediatricians and the families involved in the research project. We thank two local pediatricians' associations, the Italian Federation of Pediatricians and the Italian Confederation of Pediatricians, and their representatives, Dr. Maria Rosaria Filograna and Dr. Sergio Romano. We thank the two departments of the Local Health Service that supported the research. Specifically, the Department of Mental Health, directed by Dr. Serafino De Giorgi, and the Neuropsychiatric Unit, directed by Dr. Angelo Massagli. Finally, we thank Dr. Luigia Duma for the help during the data collection.

A.
Appendix.
The Quantitative-CHecklist for Autism in Toddler

Original version by Allison et al. (2008). Italian translation by Levante, Petrocchi, and Lecciso. Per favore rispondi alle seguenti domande sul tuo bambino scegliendo la risposta che ritieni più appropriata. Non ci sono risposte giuste o sbagliate. Ti preghiamo di rispondere a ogni domanda nella maniera che meglio descrive il tuo bambino una settimana primo o dopo il compimento dei 18 mesi di vita. Prova a rispondere se puoi a CIASCUNA domanda.

1.Il tuo bambino ti guarda quando lo chiami per nome?

- sempre
- di solito
- qualche volta
- raramente
- mai

2. Quanto è facile per te ottenere un contatto oculare con il tuo bambino?

- molto semplice
- abbastanza semplice
- abbastanza difficile
- molto difficile
- impossibile

3. Quando il tuo bambino gioca da solo, allinea gli oggetti?

- sempre
- di solito
- qualche volta
- raramente
- mai

4. Le altre persone riescono a comprendere facilmente il linguaggio del tuo bambino?

- sempre

- di solito
- qualche volta
- raramente
- mai
- il mio bambino non parla

5. Il tuo bambino punta il dito per indicare che vuole qualcosa (per esempio un giocattolo che è fuori dalla sua portata)?

- molte volte al giorno
- poche volte al giorno
- poche volte a settimana
- meno di una volta a settimana
- mai

6. Il tuo bambino punta il dito per condividere un interesse con te (per esempio indicare un evento interessante)?

- molte volte al giorno
- poche volte al giorno
- poche volte a settimana
- meno di una volta a settimana
- mai

7. Per quanto tempo l'interesse del tuo bambino può essere mantenuto su oggetti rotanti (per esempio lavatrici, ventilatori, ruote di macchine giocattolo)?

- alcune ore
- mezz'ora
- 10 minuti
- un paio di minuti
- meno di un minuto

8. Quante parole riesce a dire il tuo bambino?

- non ha ancora iniziato a parlare

- meno di 10 parole
- 10–50 parole
- 51–100 parole
- più di 100 parole

9. Il tuo bambino gioca a far finta (per esempio si prende cura di bambole, parla a un telefono giocattolo)?

- molte volte al giorno
- poche volte al giorno
- poche volte a settimana
- meno di una volta a settimana
- mai

10. Il tuo bambino segue la stessa direzione in cui tu stai guardando?

- molte volte al giorno
- poche volte al giorno
- poche volte a settimana
- meno di una volta a settimana
- mai

11. Quanto spesso il tuo bambino odora o lecca oggetti insoliti?

- molte volte al giorno
- poche volte al giorno
- poche volte a settimana
- meno di una volta a settimana
- mai

12. Il tuo bambino posa la tua mano su un oggetto quando lo vuole usare (per esempio sulla maniglia di una porta quando vuole che tu la apra, su un gioco quando vuole che tu lo azioni)?

- molte volte al giorno
- poche volte al giorno

- poche volte a settimana
- meno di una volta a settimana
- mai

13. Il tuo bambino cammina sulle punte dei piedi?

- sempre
- di solito
- qualche volta
- raramente
- mai

14. Quanto è facile per il tuo bambino adattarsi quando cambiano le sue routine o quando le cose sono fuori dal loro solito posto?

- molto semplice
- abbastanza semplice
- abbastanza difficile
- molto difficile
- impossibile

15. Se tu o qualcun'altro in famiglia siete visibilmente turbati, il tuo bambino mostra segni di volervi confortare (per esempio toccando i capelli, abbracciando)?

- sempre
- di solito
- qualche volta
- raramente
- mai

16. Il tuo bambino fa la stessa cosa più e più volte (per esempio aprire e chiudere il rubinetto, accendere e spegnere la luce, aprire e chiudere le porte)?

- molte volte al giorno
- poche volte al giorno
- poche volte a settimana

- meno di una volta a settimana
- mai

17. Descriveresti le prime parole del tuo bambino come:

- molto tipiche
- abbastanza tipiche
- un poínsolite
- molto insolite
- il mio bambino non parla

i

18. Il tuo bambino ripete le cose che sente (per esempio cose che tu dici, frasi tratte da canzoni o film, suoni)?

- molte volte al giorno
- poche volte al giorno
- poche volte a settimana
- meno di una volta a settimana
- mai

19. Il tuo bambino usa gesti semplici (per esempio fare "ciao" con la mano)?

- molte volte al giorno
- poche volte al giorno
- poche volte a settimana
- meno di una volta a settimana
- mai

20. Il tuo bambino compie movimenti insoliti con le dita vicino ai suoi occhi?

- molte volte al giorno
- poche volte al giorno
- poche volte a settimana
- meno di una volta a settimana
- mai

21. Il tuo bambino guarda spontaneamente il tuo viso per verificare la tua reazione quando affronta qualcosa di non familiare?

- sempre
- di solito
- qualche volta
- raramente
- mai

22. Per quanto tempo l'interesse del tuo bambino può essere mantenuto su uno o due oggetti?

- la maggior parte della giornata
- alcune ore
- mezz'ora
- 10 minuti
- un paio di minuti

23. Il tuo bambino usa/muove oggetti ripetutamente (per esempio pezzi di corda)?

- molte volte al giorno
- poche volte al giorno
- poche volte a settimana
- meno di una volta a settimana
- mai

24. Il tuo bambino ti sembra ipersensibile al rumore?

- sempre
- di solito
- qualche volta
- raramente
- mai

25. Il tuo bambino guarda nel vuoto senza nessun motivo apparente?

- molte volte al giorno

- poche volte al giorno
- poche volte a settimana
- meno di una volta a settimana
- mai

References

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). *The Diagnostic and Statistical Manual of Mental Disorders-5*. APA.

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

Benítez-Burraco, A. and Progovac, L. (2020). A four-stage model for language evolution under the effects of human self-domestication. *Language & Communication*, 73:1–17.

Capirci, O., Contaldo, A., Caselli, M. C., and Volterra, V. (2005). From action to language through gesture: A longitudinal perspective. *Gesture*, 5(1-2):155–177.

Chang, C. C., Wu, T. H., Chen, C. Y., Wang, J. D., and Lin, C. Y. (2014). Psychometric evaluation of the internalized stigma of mental illness scale for patients with mental illnesses: measurement invariance across time. *PLoS One*, 9(6):e98767.

Cheung, G. W. and Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-cultural Psychology*, 31(2):187–212.

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.

Devescovi, R., Monasta, L., Bin, M., Bresciani, G., Mancini, A., Carrozzi, M., and Colombi, C. (2020). A two-stage screening approach with i-tc and q-chat to identify toddlers at risk for autism spectrum disorder within the italian public health system. *Brain Sciences*, 10(3):184.

Haine-Schlagel, R., Rieth, S., Dickson, K. S., Brookman-Frazee, L., and Stahmer, A. (2020). Adapting parent engagement strategies for an evidence-based parent-mediated intervention for young children at risk for autism spectrum disorder. *Journal of Community Psychology*.

Harrop, C., McConachie, H., Emsley, R., Leadbitter, K., Green, J., Consortium, P., et al. (2014). Restricted and repetitive behaviors in autism spectrum disorders and typical development: Cross-sectional and longitudinal comparisons. *Journal of Autism and Developmental Disorders*, 44(5):1207–1219.

Heal, R. and Twycross, A. (2015). Validity and reliability in quantitative research. *Evidence-Based Nursing*, pages 66–67.

Honey, E., Leekam, S., Turner, M., and McConachie, H. (2007). Repetitive behaviour and play in typically developing children and children with autism spectrum disorders.

Journal of Autism and Developmental Disorders, 37(6):1107–1115.

Hu, L.-t. and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal*, 6(1):1–55.

Istvan, E. M., Nevill, R. E., and Mazurek, M. O. (2020). Sensory over-responsivity, repetitive behavior, and emotional functioning in boys with and without autism spectrum disorder. *Research in Autism Spectrum Disorders*, 75:101573.

Koo, T. K. and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2):155–163.

Lameira, A. R. and Call, J. (2020). Understanding language evolution: Beyond pan-centrism. *BioEssays*, 42(3):1900102.

Lecciso, F., Levante, A., Signore, F., and Petrocchi, S. (2019). Preliminary evidence of the structural validity and measurement invariance of the quantitative-checklist for autism in toddler (q-chat) on italian unselected children. *Electronic Journal of Applied Statistical Analysis*, 12(2):320–340.

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.

Lee, S. T. (2018). Testing for measurement invariance: Does your measure mean the same thing for different participants? *Association for Psychological Sciences*, 31(8).

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.

Leo, M., Carcagnì, P., Del Coco, M., Spagnolo, P., Mazzeo, P. L., Celeste, G., Distante, C., Lecciso, F., Levante, A., Rosato, A. C., et al. (2018). Towards the automatic assessment of abilities to produce facial expressions: The case study of children with asd.

Leo, M., Carcagnì, P., Distante, C., Mazzeo, P. L., Spagnolo, P., Levante, A., Petrocchi, S., and Lecciso, F. (2019). Computational analysis of deep visual data for quantifying facial expression production. *Applied Sciences*, 9(21):4542.

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 Disability*, 22:55–75.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., Bishop, S., et al. (2012). *Autism diagnostic observation schedule–2nd edition (ADOS-2)*.

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.

Marzorati, C., Mazzocco, K., Monzani, D., Pavan, F., Casiraghi, M., Spaggiari, L., Monturano, M., and Pravettoni, G. (2020). One-year quality of life trends in ealry stage lung cancer patients after lobectomy. *Frontiers in Health Psychology*, page in press.

Marzorati, C., Monzani, D., Mazzocco, K., Pavan, F., Cozzi, G., De Cobelli, O., Monturano, M., and Pravettoni, G. (2019). Predicting trajectories of recovery in prostate cancer patients undergone robot-assisted radical prostatectomy (rarp). *PloS one*, 14(4):e0214682.

Mohammadian, M., Zarafshan, H., Mohammadi, M. R., and Karimi, I. (2015). Evaluating reliability and predictive validity of the persian translation of quantitative-checklist for autism in toddlers (q-chat). *Iranian Journal of Psychiatry*, 10(1):64.

Molino, M., Ingusci, E., Signore, F., Manuti, A., Giancaspro, M. L., Russo, V., Zito, M., and Cortese, C. G. (2020). Wellbeing costs of technology use during covid-19 remote working: an investigation using the italian translation of the technostress creators scale. *Sustainability*, 12(15):5911.

Olino, T. M. (2020). Clinical applications of measurement invariance. *Journal of Personality Assessment*, 102(5):727–729.

Park, S., Won, E.-K., Lee, J. H., Yoon, S., Park, E. J., and Kim, Y. (2018). Reliability and validity of the korean translation of quantitative-checklist for autism in toddlers: a preliminary study. *Journal of the Korean Academy of Child and Adolescent Psychiatry*, 29(2):80–85.

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., Labrie, N. H., and Schulz, P. J. (2020a). 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*, 25(4):558–569.

Petrocchi, S., Levante, A., and Lecciso, F. (2020b). Systematic review of level 1 and level 2 screening tools for autism spectrum disorders in toddlers. *Brain Sciences*, 10(3):180.

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*, pages 1–146.

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.

Robins, D. L., Casagrande, K., Barton, M., Chen, C.-M. A., Dumont-Mathieu, T., and Fein, D. (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (m-chat-r/f). *Pediatrics*, 133(1):37–45.

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.

Rodgers, M., Marshall, D., Simmonds, M., Le Couteur, A., Biswas, M., Wright, K., Rai, D., Palmer, S., Stewart, L., and Hodgson, R. (2020). Interventions based on early intensive applied behaviour analysis for autistic children: a systematic review and cost-effectiveness analysis. *Health Technology Assessment*, pages 1–306.

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. (2019a). 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.

Ruta, L., Chiarotti, F., Arduino, G. M., Apicella, F., Leonardi, E., Maggio, R., Carrozza, C., Chericoni, N., Costanzo, V., Turco, N., et al. (2019b). Validation of the quantitative checklist for autism in toddlers (q-chat) in an italian clinical sample of young children with autism and other developmental disorders. *Frontiers in Psychiatry*, 10:488.

Sinai-Gavrilov, Y., Gev, T., Mor-Snir, I., Vivanti, G., and Golan, O. (2020). Integrating the early start denver model into israeli community autism spectrum disorder preschools: Effectiveness and treatment response predictors. *Autism*, 24(8):2081–2093.

Thabtah, F. and Peebles, D. (2019). Early autism screening: A comprehensive review. *International Journal of Environmental Research and Public Health*, 16(18):3502.

Tregay, J., Gilmour, J., and Charman, T. (2009). Childhood rituals and executive functions. *British Journal of Developmental Psychology*, 27(2):283–296.

Tyrell, F. A., Yates, T. M., Widaman, K. F., Reynolds, C. A., and Fabricius, W. V. (2019). Data harmonization: Establishing measurement invariance across different assessments of the same construct across adolescence. *Journal of Clinical Child & Adolescent Psychology*, 48(4):555–567.

Volterra, V. (1981). Gestures, signs, & words at two years: when does communication become language? *Sign Language Studies*, (33):351–362.

Widaman, K. F., Ferrer, E., and Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, 4(1):10–18.