UvA-DARE (Digital Academic Repository)

How to measure camouflaging? A conceptual replication of the validation of the Camouflaging Autistic Traits Questionnaire in Dutch adults

van der Putten, W.J.; Agelink van Rentergem, J.A.; Radhoe, T.A.; Torenvliet, C.; Groenman, A.P.; Mol, A.J.J.; Geurts, H.M.

Link to publication

How to measure camouflaging? A conceptual replication of the validation of the Camouflaging Autistic Traits Questionnaire in Dutch adults

Wikke J. van der Putten [a,b,*], Joost A. Agelink van Rentergem [b], Tulsi A. Radhoe [b], Carolien Torenvliet [b], Annabeth P. Groenman [b], Audrey J.J. Mol [a], Hilde M. Geurts [a,b]

[a] *Leo Kannerhuis, autism outpatient clinic (Youz/Parnassia Group), Derkinderenstraat 10–18, 1062 DB Amsterdam, the Netherlands*
[b] *Dutch Autism & ADHD Research Center, Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129, 1001 NK Amsterdam, the Netherlands*

A B S T R A C T

Background: Camouflaging behavior is often defined as using strategies to hide autistic characteristics. In this study, we investigated how to measure camouflaging behavior by performing a conceptual replication of the original study of the Camouflaging Autistic Traits Questionnaire (CAT-Q) and testing whether the self-reported camouflaging behavior is measuring the same construct as the second most used manner to measure camouflaging behavior, the discrepancy method.
Method: In total, 674 individuals (356 autistic) aged 30–92 years, filled out the Dutch translation of the CAT-Q (CAT-Q-NL) and the Autism Spectrum Quotient (AQ). In 90 autistic adults the Autism Diagnostic Observation Schedule (ADOS-2) was administered. We executed preregistered analyses (AsPredicted #37800) to investigate the factor structure, measurement invariance, internal consistency, convergent validity and group differences.
Results: Our analyses showed that the original three-factor structure had an acceptable fit and internal consistency ranged from sufficient to good. However, there was no measurement invariance between autistic and non-autistic individuals and correlations between CAT-Q-NL-scores and the discrepancy between AQ and ADOS-2 varied between low to mediocre (r = 0.04 to.28).
Conclusions: The CAT-Q-NL can be used to measure camouflaging between and within autistic adults, but not between autistic and non-autistic adults and its convergent validity is limited. Despite these caveats, the CAT-Q-NL can serve as a useful addition to the clinical assessment toolbox because gaining insight in the level of camouflaging of autistic adults may help provide better mental health care. However, more research is needed into how to optimally measure the camouflaging construct.

* Correspondence to: Leo Kannerhuis, autism outpatient clinic (Youz/Parnassia Group), Overschiestraat 57, 1062 HN Amsterdam, the Netherlands.
E-mail addresses: w.vanderputten@leokannerhuis.nl, w.vanderputten@leokannerhuis.nl (W.J. van der Putten).

1. Introduction

Camouflaging behavior[1] has recently gained attention in scientific research related to autism[2] (Cook et al., 2021; Libsack et al., 2021). Camouflaging behavior has been described as the conscious or unconscious use of strategies to hide one's autism characteristics (Hull et al., 2017). Autism is often defined as experiencing difficulties in social communication and interaction, and the presence of restricted, repetitive patterns of behavior, interests or activities (American Psychiatric Association, 2013). Thus, when people camouflage, they attempt to minimize, alter, or otherwise change the presentation of their autistic behavior to pass as non-autistic (Libsack et al., 2021). Most research is currently focusing on the causes and consequences of camouflaging (Cassidy et al., 2019; Livingston et al., 2019), but it is still not evident how to best measure this construct (Cook et al., 2021; Libsack et al., 2021; Williams, 2022). Therefore, we focus on different methods that are most often used to measure camouflaging.

Developing a valid instrument to measure camouflaging behavior is important for clinical practice and scientific research. In qualitative studies, autistic adults reported camouflaging as a cause for negative consequences, such as missed or late diagnoses and experiencing mental health difficulties (Bargiela et al., 2016; Hull et al., 2017; Livingston, Shah et al., 2019). Several cross-sectional studies reported that camouflaging behavior is associated with depression, anxiety, suicidality and reduced wellbeing. However, these findings have not yet been replicated in longitudinal studies and therefore, we cannot yet draw conclusions about causality (Cook et al., 2021). (Semi-)structured interviews have been conducted with autistic teenagers and adults in order to gain more insight in, and deeper understanding of camouflaging behavior (Bargiela et al., 2016; Livingston et al., 2019; Tierney et al., 2016). Interviews with late diagnosed autistic females show they adopted specific strategies to hide their autistic traits based on television, magazines and books about body language (Bargiela et al., 2016). Also, autistic female adolescents describe copying behaviors of others (Tierney et al., 2016). Some mention that they are even taking on a certain persona when they are in social situations (Bargiela et al., 2016; Tierney et al., 2016). Aside from consciously learned strategies, autistic adults also report unconsciously mimicking speech patterns and body language, for example unintentionally copying specific accents (Bargiela et al., 2016). It has been hypothesized that autistic adults camouflage as a conscious or unconscious reaction to being stigmatized or marginalized (Pearson & Rose, 2021).

In the past years several approaches of measuring camouflaging behavior have been used. One approach is to operationalize camouflaging behaviors as the discrepancy between "internal autistic status" and external presentation of autism traits (Lai et al., 2016). An extensive overview of the used forms of the discrepancy approach has recently been provided by Libsack and colleagues (2021). For example, many studies operationalized this discrepancy method by comparing observable autism behavior (external presentation) based on the Autism Diagnostic Observation Scale (ADOS-2; Lord et al., 2000) or Autism Diagnostic Interview-Revised (ADI-R; Lord et al., 1994) to internal autistic status measured using self- or parent-reported autism traits (Lai et al., 2016; Ratto et al., 2018). Another method of the discrepancy approach is to measure someone's internal autism status using its assumed underlying cognition, e.g., Theory of Mind (Livingston, Colvert, et al., 2019). However, the validity of measuring camouflaging using these two methods of the discrepancy approach has recently been debated (Fombonne, 2020; Lai et al., 2021; Williams, 2022). Therefore, to further investigate camouflaging behavior, a specific instrument is necessary to directly investigate this construct.

The first validated self-report questionnaire that measures social camouflaging behavior is the CAT-Q (Hull et al., 2019). This questionnaire consists of 25 items describing different types of camouflaging behavior, based on experiences of autistic adults. Exploratory Factor Analyses revealed a three-factor structure consisting of the following factors: 1) compensation (actively compensating for difficulties in social situations), 2) masking (hiding one's autistic characteristics) and 3) assimilation (trying to fit in with others in social situations). Equivalent factor loadings, intercepts and residuals were found for the CAT-Q in autistic and non-autistic males and females and therefore, the CAT-Q can be used to compare camouflaging behavior across groups (Hull et al., 2019). Also, they found that scores on the CAT-Q correlated with self-reported wellbeing, social and general anxiety, depression and number of autism traits (Hull et al., 2019). Therefore, the validity and reliability of the CAT-Q were reported sufficient to good and the CAT-Q seems to be a promising instrument for both research and clinical practice. The CAT-Q is one of the most widely used standardized instruments to date (Cook et al., 2021; Libsack et al., 2021). However, several issues have been raised about some psychometric properties of the CAT-Q (Fombonne, 2020; Williams, 2022). Therefore, for our study we translated the CAT-Q to Dutch (CAT-Q-NL) to be able to focus on further examining these psychometric properties and with that to also investigate the validity of this Dutch version.

There are some important issues to address. First, as the CAT-Q-NL will be used to investigate group differences in camouflaging behavior, it is important to ensure that the same construct is measured in all groups by establishing measurement invariance (Wicherts & Dolan, 2010). It is essential to investigate measurement invariance in autistic and non-autistic males and females as the differences in camouflaging behavior are of interest and scores on the CAT-Q have been shown to differ across these groups (Hull et al., 2020). That is, autistic females showed more assimilation and masking behavior, than autistic males, while non-autistic males showed more compensation behavior than non-autistic females. Measurement invariance (which is similar to differential item functioning [DIF] (Takane & de Leeuw, 1987)) indicates whether differences in scores actually represent different levels of camouflaging, or whether the construct is measured differently in both groups. We will test whether this finding can be replicated for the CAT-Q-NL.

Secondly, an important criticism on the CAT-Q is that its convergent validity has been assessed insufficiently (Fombonne, 2020). In

---

[1] In this paper we use the term camouflaging behavior, but different terms are currently used, such as masking, compensating, passing as non-autistic, and adaptive morphing (Lawson, 2020; Libsack et al., 2021).

[2] In this paper, the term "autism" instead of "autism spectrum disorder (ASD" and identity-first language instead of person-first language is used, as this is preferred by most autistic adults (Kenny et al., 2016).

recent work, convergent validity of the CAT-Q was assessed by correlating scores on the CAT-Q with self-reported wellbeing, social and general anxiety, depression and number of autism traits (Hull et al., 2019). While several studies indeed indicate that there is a relationship between camouflaging behavior and poorer mental health (Cook et al., 2021), this correlation cannot be seen as evidence for convergent validity. That is, based on these results, we cannot distinguish whether the CAT-Q measures camouflaging behavior, autism traits, mental health or social anxiety (Fombonne, 2020). The discrepancy method could be a better alternative to assess the convergent validity of the CAT-Q because the discrepancy method is used the specifically measure camouflage behavior even though it is, as aforementioned, not without caveats. Therefore, in our study we will compare the CAT-Q-NL to the discrepancy approach as an operationalization of camouflaging behavior.

To sum up, in this preregistered study we aim to investigate: 1) the factor structure of the CAT-Q-NL, 2) measurement invariance for autistic and non-autistic males and females, 3) internal reliability of the CAT-Q-NL; 4) convergent validity by comparing the CAT-Q-NL with the discrepancy between Autism-Spectrum Quotient (AQ; Baron-Cohen et al., 2001) and total ADOS-2 score and 5) group differences between autistic and non-autistic males and females. By doing so we aim to replicate results of Hull et al. (2019) regarding the original CAT-Q and to determine whether the CAT-Q(-NL) can indeed be seen as a psychometrically sound addition to the clinically available assessment toolbox for autistic adults.

2. Methods

2.1. Participants

A total of 674 (356 autistic and 318 non-autistic) adults participated in this study. Autistic individuals were recruited through mental health institutions across the Netherlands, by means of advertisement on client organization websites, newsletters, and social media (i.e., Twitter and LinkedIn). Non-autistic participants were recruited through social media and personal networks of researchers and participants. Autistic and non-autistic participants of a previous study with an age of 30 or higher, were also invited to participate in this study (Geurts et al., 2021). Group characteristics are shown in Table 1.

Inclusion criteria for all participants were: 1) no self-reported intellectual disability and 2) sufficient understanding of the Dutch language (at least one Dutch parent, Dutch was spoken in their family or when participants are able to fill in more than 90% of the questionnaires). Individuals in the autism (AUT) group reported a clinical autism diagnosis, including year and mental health institution of diagnosis. Additional inclusion criteria for the non-autistic comparison group (COMP) were: 1) no present/past autism or Attention Deficit Hyperactivity Disorder (ADHD) diagnosis, 2) no first-degree relatives family members with an autism or ADHD-diagnosis and 3) AQ $\leq$ 32 and DSM-IV criteria during childhood and adulthood for inattention and hyperactivity/impulsivity < 6 based on the ADHD-Rating Scale (ADHD-SR; Kooij et al., 2005)). Additional inclusion criteria for the autism subsample that was invited for a face-to-face session are 1) estimated IQ > 70, measured using two subtests (Vocabulary and Matrix reasoning) of the Wechsler Adult Intelligence Scale IV-NL (WAIS-IV-NL; Wechsler, 2012; Wechsler, 2008), 2), no history of neurological disorders (e.g., epilepsy, stroke, multiple sclerosis), schizophrenia or having experienced more than one psychosis and 3) no current alcohol or drugs dependency.

Table 1
Characteristics of the AUT and COMP group and AUT subsample.

	Total	AUT group	AUT subsample	COMP group	Test statistic (p-value)[a]
N	674	356	90	318	–
Biological sex (Male/Female)[b]	359/315	187/169	60/30	172/146	$\chi^2 = 0.16$ (0.069)
Mean age	54.4	52.3	54.2	56.7	$t = -4.21$ (<0.001)
(SD; range)	(13.5; 30–92)	(12.5; 30–84)	(14.2; 30–84)	(14.1; 30–92)	
Mean age autism diagnosis (SD; range)	–	45.99	47.6	–	–
		(13.2; 4–80)	(15.0; 8–80)		
Mean AQ total score	24.6	34.3	35.0	13.7	$t = 39.20$ (<0.001)
(SD; range)	(12.4; 2–48)	(7.5; 10–48)	(7.2; 14–48)	(5.9; 2–31)	
Mean ADOS-2 total score (SD; range)	–	–	11.2	–	–
			(3.7; 2–19)		
Mean estimated IQ	114.2[c]	115.0[c]	114.7	113.3[c]	$t = 0.92$ (0.36)
(SD; range)	(16.4; 66–147)	(16.0; 66–147)	(15.5; 81–147)	(16.9; 69–144)	

Note. ADOS-2 = Autism Diagnostic Observation Scale; AQ = Autism-spectrum Quotient; AUT = Autism; COMP = Comparison; IQ = Intelligence Quotient.

[a] Comparison of the AUT and COMP group.

[b] In the AUT group, 2 individuals reported "other" as their sex. We did not include these individuals in the study as the sample size was too low to include as a separate group.

[c] Results are only available for a subsample of this group.

2.2. Materials

2.2.1. Dutch translation of the Camouflaging Autistic Traits Questionnaire (CAT-Q-NL)

The CAT-Q-NL is an unpublished two-way translation by Agelink van Rentergem et al. (2018) of the CAT-Q (Hull et al., 2019) which is a self-report questionnaire to measure camouflaging behavior. With permission of the authors, the CAT-Q was translated into Dutch. The CAT-Q-NL was translated back into English by an independent translator and this translation was approved by the original authors of the CAT-Q. The CAT-Q-NL consists of 25 items describing different types of camouflaging behavior. Participants indicated on a 7-point Likert scale whether they "strongly disagree" (1) to "strongly agree" (7) with each statement. The total score ranges from 25 to 175 and a higher score indicates more camouflaging behavior.

2.2.2. Autism Spectrum Quotient (AQ; Baron-Cohen et al., 2001; Hoekstra et al., 2008)

The AQ is a self-report questionnaire that measures autism traits and consists of 50 items scored on a 4-point Likert scale (1 "definitely agree" to 4 "definitely disagree"). Items are rescored to a 0 or 1 following the algorithm of the AQ, with 1 indicating autistic-like behavior, resulting in a total score ranging from 0 to 50. Psychometric properties of the AQ are satisfactory (Baron-Cohen et al., 2001; Hoekstra et al., 2008).

2.2.3. Module 4 of the Autism Diagnostic Observation Scale, version 2 (ADOS-2; (Bildt et al., 2013; Lord et al., 2012)

The ADOS-2 is a semi-structured, standardized assessment measuring social interactions, communication, and stereotypical behaviors. The ADOS-2 was administered by trained researchers (CT and TAR). The Module 4 revised algorithm has a sensitivity and specificity of 80% (Hus & Lord, 2014) and was used to calculate a total ADOS-2-score ranging from 0 to 22.

2.3. Procedure

The present study is part of a large ongoing longitudinal project investigating "Autism & Aging" (full procedure is described in Geurts et al., 2021). The study was approved by the ethical commission of the University of Amsterdam (2018-BC-9285). All participants gave written consent for participation in the study. Participants completed two series of questionnaires, including the CAT-Q-NL, AQ (Baron-Cohen et al., 2001; Hoekstra et al., 2008) and ADHD-SR (Kooij et al., 2005). Hereafter, a subsample of participants was invited to face-to-face sessions during which, among others, the ADOS-2 (Bildt et al., 2013; Lord et al., 2012) and subtests of the WAIS-IV-NL (Wechsler, 2012) were administered. Participants received compensation for travel expenses and a small monetary reward for participating in the study.

2.4. Community involvement

Four (older) autistic adults were involved in the "Autism & Aging" project (Geurts et al., 2021), of which the present study is part of. This group advised, among others, on recruitment of participants, study design, information letters and interpretation of results. Researchers meet the group about four times each year over a period of five years and the group is paid for their contributions.

2.5. Statistical analyses

Analyses were executed with Rstudio (RStudio Team, 2020), using the "lavaan package" (Rosseel, 2012). The study design and statistical analyses plan was preregistered on AsPredicted.org (#37800) . The total sample (Sample AB, $N_{AB}= 674$) was randomly split in Sample A ($N_A= 335$) and Sample B ($N_B= 339$). Characteristics of Sample A and B are described in Table S2.

2.5.1. Confirmatory factor analysis (CFA)

We investigated whether we could replicate the factor structure found by Hull et al. (2019). Put differently we tested whether the three factor structure was a good fit for the CAT-Q-NL, using CFA with maximum likelihood estimator in Sample AB. We made holistic judgements based on following criteria: a mediocre fit is represented by a Root Mean Square Error of Approximation (RMSEA) < 0.08, a Good Fit Index (GFI) > 0.90 and a Standardized Root Mean Square Residual (SRMR) < 0.10. A good fit is represented by RMSEA< 0.06, GFI> 0.95, and SRMR< 0.08 (Brown, 2015). We reported the chi-square results, however, we do not use this fit index to make a judgement of the fit, as this is not a reliable index for large samples (Sass, 2011).

2.5.2. Measurement invariance analyses

When the results of the CFA indicated a mediocre to good fit, we investigated whether, like in Hull et al. (2019) the factor structure is similar across autistic and non-autistic males and females. We used the same approach as in the original paper and, therefore, we investigated configural, metric, scalar and residual invariance. We reported Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and difference in Comparative Fit Index (ΔCFI), for which lower is better. If ΔCFI ≤ 0.01, this indicates measurement invariance (Sass, 2011).

2.5.3. Exploratory factor analysis (EFA)

As preregistered, if measurement invariance would not be found for the original factor structure (see results below), the factor structure of the CAT-Q-NL would be investigated using an EFA in Sample A. The number of factors to retain is based on parallel analysis

(Hayton et al., 2004) and a visual inspection of the scree plot. We removed items with high loadings on more than one factor (cross-loadings <0.20) and items with small loadings on all factors (factor loadings <0.30) (Brown, 2015). After executing the EFA, we investigated whether the alternative factor structure could be replicated in Sample B and whether we would find measurement invariance for this factor structure.

2.5.4. Reliability

We investigated whether we would replicate the sufficient to high internal consistency found by Hull et al. (2019) for all factors. We calculated Cronbach's α for subscales and total scores, separately in the AUT and COMP group. Guidelines of the Dutch Committee on Tests and Testing (COTAN) were followed (Evers et al., 2009) for the interpretation: $r \geq 0.80$ is good, $0.70 \leq r < 0.80$ is sufficient and $r < 0.70$ is insufficient.

2.5.5. Convergent validity

Convergent validity was investigated by comparing the CAT-Q-NL-scores to the discrepancy measure. In the study of Lai and colleagues (2016), the discrepancy scores were calculated by a combination between the discrepancy between the ADOS-2 and the AQ, and the discrepancy between the ADOS and a social cognition task: Reading the Mind in the Eyes. In our study, we focus on the discrepancy between the ADOS-2 and the AQ, similar to Schuck et al. (2019). The discrepancy measure is calculated by subtracting the standardized total ADOS-2 algorithm score from the standardized AQ-score. A high score on this discrepancy measure represents more camouflaging behavior. A high correlation ($r = 0.50$) is considered evidence for good convergent validity and a moderate correlation ($r = 0.30$) as sufficient.

2.5.6. Group differences

Due to our measurement invariance results (see Results), we deviated from the preregistered analysis plan and only investigated group differences between males and females separate for the AUT and COMP group. In addition, we calculated 95%-confidence intervals in order to interpret whether individual camouflaging scores are higher or lower than most individuals of a certain group. Group comparisons were checked using the "Statcheck" package (Nuijten, 2018).

3. Results

Please see Fig. 1 for a summary of our results and every single statistical step taken.

3.1. Step 1. Can we replicate the original CAT-Q factor structure found by Hull and colleagues (2019)?

Using a CFA, we investigated whether we could replicate the three-factor structure found by Hull et al. (2019) in Sample AB. The results, varied slightly across fit measures (see Table 2). We found a good fit based on the SRMR, a significant chi-square value and a mediocre fit based on the GFI. The structure did not fit based on the RMSEA. Since two out of three fit indices showed a moderate to good fit, we concluded that we could replicate Hull et al. (2019) as the original three-factor structure sufficiently explained the data.
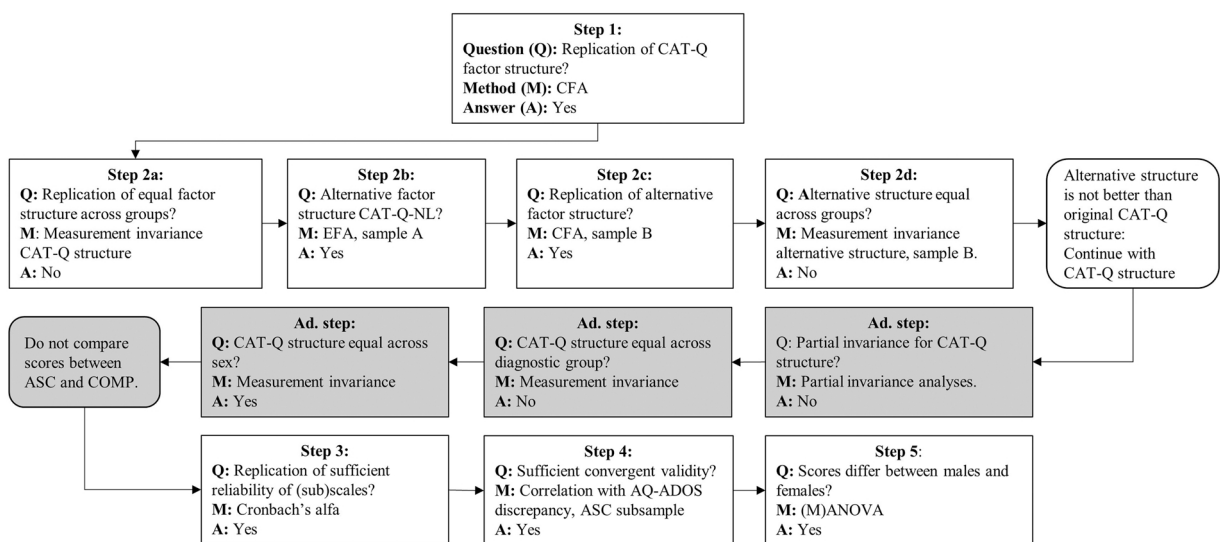

Fig. 1. Summary of statistical steps made in the results section. Note. When the sample is not specified, the analyses were executed in Sample AB. Steps in grey are exploratory. Step 3 and 5 were slightly altered from the original preregistration due to the lack of measurement invariance.

Table 2
Test statistics of the CFA testing the replication of the alternative factor structure on Sample B.

	χ^2	Df	RMSEA (90% CI)	GFI	SRMR
Original factor structure Sample AB	1665.03 * **	272	0.087 (0.083-0.091)	0.923	0.070
Alternative factor structure Sample B[a]	334.124 * **	116	0.074 (0.065-0.084)	0.960	0.060

Note. [a] This is the factor structure that resulted from the EFA in Sample A.
Df= degrees of freedom; * ** = $p < .001$; RMSEA = Root Mean Square Error of
Approximation; 90% CI = 90% Confidence Interval; GFI = Good Fit Index; SRMR = Standardized Root Mean Square Residual

3.2. Step 2a. Can we replicate that the factor structure is equal across autistic and non-autistic males and females?

Measurement invariance analyses were executed across autistic and non-autistic males and females to investigate whether the factor structure, loadings, intercepts, and residual variances were equal across all four groups. The results of the CFA separate for each group are shown in Table S1 in the supplementary materials. The results in Table 3 show that metric, scalar, and residual invariances were violated because ΔCFI > 0.01. Factor loadings, intercepts and variances were not equal across all groups. Thus, we could not replicate Hull et al. (2019) as the CAT-Q-NL does not measure camouflaging behavior in the same manner in autistic and non-autistic males and females. Despite sufficient fit of the original factor structure we investigated, as preregistered, whether an alternative factor model would not encounter problems of measurement invariance.

3.3. Step 2b, c, & d. What is the factor structure of the CAT-Q-NL investigated by means of EFA and is this factor structure equal across autistic and non-autistic males and females?

Using EFA, we investigated if the problems of measurement invariance would disappear with an alternative factor structure. Scree plot and parallel analysis (see: statistical analyses) indicated that three factors best explained the data. Therefore, we ran the EFA with three factors to investigate the factor structure of the CAT-Q-NL. Factor loadings for all 25 items are shown in Table S3. We removed items with cross loadings < 0.20 (items 3, 4, 5, 8, 9, 11, 14) or factor loadings < 0.30 (item 12). There did not seem to be a specific similarity between the items that were removed based on the EFA, however we could not empirically verify this. This resulted in an alternative factor structure, consisting of 16 items. Factor 1 consisted of items 7, 10, 13, 16, 19, 22 and 25. All items originally belonged to the assimilation factor. Factor 2 consisted of item 1, 2, 6, 15, 18, 21 and 24, which mainly belonged to the original masking factor (all except item 1). Factor 3 consisted of item 17, 20 and 23, these are all items that originally belonged to the compensation factor.

Using CFA, we tested whether the alternative factor structure found in Sample A, could be replicated in Sample B. Results are shown in Table 3, and indicate a good fit based on the chi-square, GFI, and SRMR, and a mediocre fit based on the RMSEA. Overall, we conclude that the alternative factor structure fitted the data well. Again, we investigated whether the alternative factor structure was equal across autistic and non-autistic males and females, using measurement invariance analyses in Sample B. Like with the original factor structure, we did not find metric, scalar and residual invariance (see Table S4). Therefore, using an alternative factor structure does not solve problems of measurement invariance. Therefore, we further investigate the original factor structure.

3.4. Exploratory analyses investigating measurement invariance violations

Because measurement invariance was violated for the original and alternative factor structure, we explored if we could reach partial measurement invariance. Finding partial invariance would imply there are only a few items that differ across groups. The results showed that we could not reach partial invariance and they are described in Table S5 (see Supplementary Material). Results indicate that violations of measurement invariance were not due to a few items, but that it was a pattern across most items. To gain insight in the specific differences in measurement invariances across four groups, we separately explored measurement invariance for diagnostic group and sex, shown in Table 4. These results show that factor loadings, intercepts, and residual variances differ across autistic and non-autistic adults, but are similar across males and females. Based on results of all measurement invariance analyses, we conclude that we cannot compare scores on the CAT-Q-NL between autistic and non-autistic adults.

Table 3
Results of the measurement invariance analyses for the original factor structure in Sample AB.

	χ^2	$\Delta \chi^2$	Df	CFI	ΔCFI	AIC	BIC
Configural invariance	2592.96	–	1088	0.782	–	59452.203	60860.331
Metric invariance	2807.42 * **	214.46	1154	0.761	0.021	59534.657	60644.911
Scalar invariance	3078.90 * **	271.48	1220	0.730	0.031	59129.144	59939.916
Residual invariance	3793.54 * **	714.64	1295	0.638	0.116	60238.775	60712.664

Note. Df= degrees of freedom; * ** = $p < .001$; CFI = Comparative fit index; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

Table 4

Results of measurement invariance analyses separately in group and sex for the original CAT-Q factor structure in Sample AB.

	χ^2	$\Delta \chi^2$	DF	CFI	ΔCFI	AIC	BIC
Group							
Configural invariance	1944.37	–	544	0.792	–	59466.84	60170.91
Metric invariance	2065.84 * **	121.47	566	0.777	0.015	59544.32	60149.09
Scalar invariance	2237.74 * **	171.90	588	0.754	0.023	59131.13	59635.61
Residual invariance	2837.89 * **	600.15	613	0.669	0.085	60222.37	60615.02
Sex							
Configural invariance	1996.24	–	544	0.842	–	60711.11	61415.18
Metric invariance	2051.61 * **	55.37	566	0.838	0.004	60722.48	61327.26
Scalar invariance	2099.62 * *	48.01	588	0.834	0.004	60181.88	60686.36
Residual invariance	2156.18 * **	56.56	613	0.832	0.002	60733.05	61125.70

Note. Df= degrees of freedom; * * = $p < .01$; * ** = $p < .001$; CFI = Comparative Fit Index; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

3.5. Step 3. Is the CAT-Q-NL a reliable instrument?

As shown in Table 5, we found high internal consistency in the autism group for the total score and subscales based on the factor structure of the original CAT-Q with Cronbach's α ranging from.80 to.93. In the COMP group, we found sufficient to high internal consistency, with Cronbach's α ranging from.75 to.87. Thus, we replicated the reliability of the total and subscale scores found in the original CAT-Q (Hull et al., 2019).

3.6. Step 4. What is the convergent validity of the CAT-Q-NL?

In the autism subsample, for whom apart from the AQ also the ADOS-2 information was available (N = 90), we investigated the correlation between the discrepancy measure and the CAT-Q-NL based on the factor structure of the original CAT-Q. The autism subsample and the total autism group did not differ on total CAT-Q-NL score (M_{AUT} = 98.05, SD_{AUT} = 26.31; $M_{SUBSAMPLE}$ = 95.44, $SD_{SUBSAMPLE}$ = 27.96; $t(142) = 0.82$, $p = .41$) and total AQ score (M_{AUT} = 34.36, SD_{AUT} =7.54; $M_{SUBSAMPLE}$ = 34.83, $SD_{SUBSAMPLE}$ = 7.37; $t(151) = -0.54$, $p = .59$). Correlation coefficients varied between low, with the masking scale, to mediocre, with the assimilation scale, see Table 5. Only the correlation with assimilation significantly deviated from 0. This indicates that the CAT-Q-NL and discrepancy measure do not asses the same underlying construct.

3.7. Step 5. Do males and females in the AUT and COMP group score differently on the CAT-Q-NL?

We compared scores on subscales and total scores of the CAT-Q-NL, based on the factor structure of the original CAT-Q, between males and females separately for the AUT and the COMP group, since measurement invariance was found based on sex but not on diagnosis. The distribution of total CAT-Q-NL-scores of the AUT and COMP group are shown in Fig. S1. In the AUT group, we found that females (M= 101.3, SD= 27.4) had a significantly higher total CAT-Q-NL-score than males (M = 94.9, SD = 25.1), ($F(1, 349) = 5.29$, $p < .05$). No significant effect of sex was found on any of the subscale scores in the AUT group, $F(3, 347) = 2.05$, $p = .11$. In the COMP group we found that males (M = 68.3, SD = 16.8) had a significantly higher total CAT-Q-NL-score compared to females (M = 63.1, SD = 18.6), $F(1, 315) = 6.95$, $p < .01$. There was a significant effect of sex on subscales of the CAT-Q-NL in the COMP group, $F(3, 313) = 3.25$, $p = .043$. Males scored higher on the masking ($F(1, 315) = 8.07$, $p = .005$) and assimilation subscale ($F(1, 315) = 4.14$, $p < .05$). No difference was found on the compensation subscale ($F(1, 315) = 1.37$, $p = .24$). Sex differences in both groups on subscales of the CAT-Q-NL are shown in Fig. 2. In addition, we calculated 95% confidence intervals for the separate subscales and total scores, see Table S6.

4. Discussion

The aim of the current study was to gain more insight in measuring the construct of camouflaging. We focused on two different methods that are often used to measure camouflaging, a self-report questionnaire and the discrepancy measure. We first performed a conceptual replication of the original CAT-Q study (Hull et al., 2019) through which we investigated the psychometric properties of the CAT-Q-NL. We replicated that the three-factor structure of the CAT-Q, also fits the CAT-Q-NL, with sufficient to high internal

Table 5

Cronbach's α and correlation coefficients between the discrepancy approach and the total score and subscales of the CAT-Q-NL.

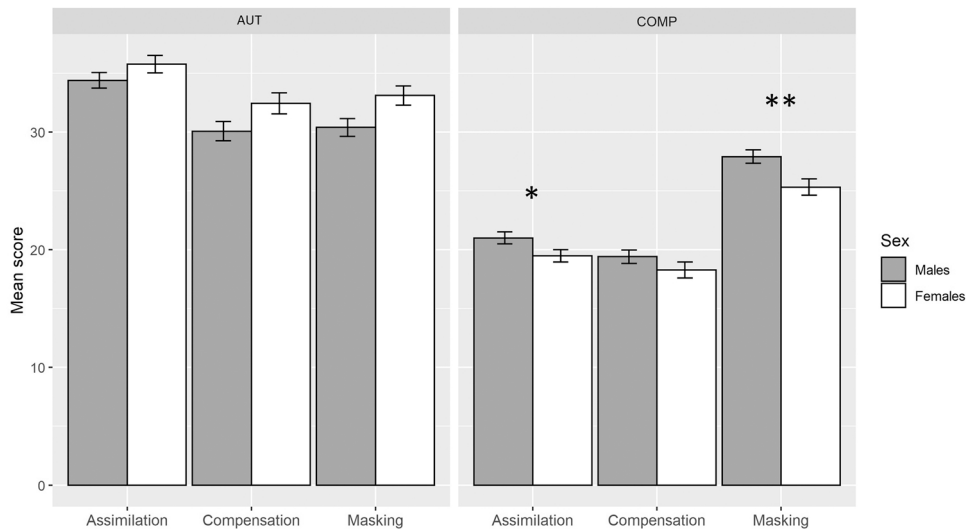|	Cronbach's α AUT group	Cronbach's α COMP group	*r* discrepancy measure AUT group (*p*-value)
Compensation	0.85	0.83	0.15 (0.16)
Masking	0.85	0.77	0.04 (0.68)
Assimilation	0.80	0.75	0.28 (0.006)
Total score	0.93	0.87	0.17 (0.10)

Fig. 2. Mean scores on subscales of the CAT-Q-NL separately for autistic (left panel) and non-autistic (right panel) males (in grey) and females (in white), with error bars representing standard errors. Asterisks indicate significant differences (*<0.05, **<0.01). The assimilation and masking scale ranges from 8 to 56 and the compensation from 9 to 63. *Note.* Due to the observed measurement invariance no direct comparison may be made between autistic and non-autistic adults.

consistency. However, we did not find measurement invariance for autistic and non-autistic adults, which indicates that the CAT-Q-NL measures camouflaging behavior differently in these groups. At the same time, we did observe measurement invariance across sex. Therefore, we cannot meaningfully compare camouflaging behavior between autistic and non-autistic adults, while we can use the CAT-Q-NL to compare camouflaging behavior between autistic males and females, and between non-autistic males and females. Thus, while we replicated the original CAT-Q factor structure and reliability findings, the major difference is that we did not observe measurement invariance. Second, we tested the convergent validity of the CAT-Q-NL. We found that self-reported camouflaging behavior is tapping into a different underlying construct than the discrepancy approach. We argue that the CAT-Q-NL can be used to gain insight in the level of camouflaging behavior between and within autistic individuals, but convergent validity and measurement invariance should be further investigated as the optimal way to measure camouflaging is not evident.

The confirmatory factor analysis results show a sufficient fit of the original three-factor structure of the CAT-Q in the CAT-Q-NL. Also, we found sufficient to high internal consistency for the CAT-Q-NL and its subscales. These results are in line with findings of Hull et al. (2019) and confirm that we can distinguish three subtypes within camouflaging behavior: compensation, masking, and assimilation. Comparisons of the levels of camouflaging behavior between autistic and non-autistic males and females showed similar trends in group differences as found by Hull et al. (2020). On average, autistic females scored higher than autistic males on all scales, but only total scores meaningfully differed. For non-autistic individuals, males scored higher, with meaningful differences on masking, assimilation and total camouflaging behavior. The subscales of the CAT-Q-NL can be helpful when they are used in clinical practice and scientific research, as they can help to investigate which types of, and what reasons for camouflaging behavior are associated with positive and negative consequences.

In contrast to findings of Hull et al. (2019), we found that factor loadings, intercepts and residual variances differed between autistic and non-autistic individuals. In such situations, it is recommended not to compare groups because comparisons are not meaningful (Putnick & Bornstein, 2016). That is, scores and differences in scores do not mean the same in both groups and therefore, comparing scores on the CAT-Q-NL between autistic and non-autistic adults would be like comparing apples and oranges. This lack of measurement invariance could not be solved by using an alternative or partial factor structure of the CAT-Q. In addition, the better fit of the alternative shorter questionnaire did not outweigh the advantages of being able to use the CAT-Q-NL in a similar manner as it is used in the international literature. One could argue that differences in measurement invariance between the current study and the original CAT-Q study (Hull et al., 2019) are due to the translation of the CAT-Q-NL. It is possible, that the Dutch items function differently or that the interpretation of certain questions is culturally determined. We followed proper procedures to translate the CAT-Q by using a two-way translation procedure and checking whether the back-translation was in agreement with respect to the original intent of the question. However, while we checked whether items needed a cultural adaptation, we did not systematically study this. Also, subtle differences between the two highly similar study designs might have been of influence. That is, the average age of our participants was higher than of the participants in the study by Hull et al. (2019) (54.4 vs 36.0 years). Although we do not expect that a lack in measurement invariance only exists in people aged 30 years and older, the way in which someone camouflages could differ based on lifespan development, age of diagnosis, or social experiences. Furthermore, we also recruited participants through mental health institutions and not solely via social media, which may have resulted in a more diverse and representative autism group.

While each of the above might be the case, the observed lack of measurement invariance does actually fit neatly with the definition of camouflaging behavior as camouflaging has specifically been defined as hiding one's autistic traits (Hull et al., 2017). As

non-autistic individuals have fewer autistic traits, it could well be that it is different for them to hide these specific traits. Moreover, the CAT-Q was developed based on experiences of camouflaging strategies reported by autistic individuals (Hull et al., 2019). This might make it more likely that the CAT-Q items resonate better with autistic individuals as compared to non-autistic individuals. However, there is debate whether or not camouflaging is actually specific and unique for autistic individuals (Fombonne, 2020; Lai et al., 2021) as individuals with other mental or physical difficulties may also experience stigma and camouflage their difficulties. While it is likely that various groups of people will camouflage in specific contexts, the intensity and specific camouflaging strategies may well differ between various groups. Our findings do suggest that indeed autistic adults have a specific way of camouflaging. This is in line with the recently proposed conceptualization of camouflaging as a type of impression management in which there is overlap between autistic and non-autistic adults, but that there are also unique differences in the experiences of autistic adults (Ai et al., 2022). Because the measurement invariance findings differ between the CAT-Q and CAT-Q-NL, further research could shed more light on whether this lack of measurement invariance is specific for Dutch adults or for this age group. Replicating this study in a broad age range in a different English speaking country and investigating measurement invariance in more detail, could help us gain more insight in how to measure the camouflaging construct.

In the present study, we compared self-reported camouflaging to the discrepancy approach in order to investigate convergent validity. Low correlation coefficients between the CAT-Q-NL and the discrepancy approach imply that we do not measure the same underlying construct with these methods. Where the CAT-Q-NL measures the extent to which someone tries to camouflage, it may not measure whether it also results in a change of observable behavior. For example, someone might report numerous autistic traits (resulting in a high AQ-score) and mention using multiple camouflaging strategies (resulting in a high CAT-Q-NL score). Despite trying to use these strategies, these persons may still not meet the demands from the environment (resulting in a high ADOS-2 score). This difference between an observable and self-report measure may be comparable to a difference that is found between subjective and objective measures of cognitive complaints: while both measures can be informative, they often do not correlate to a high degree (Groenman et al., 2022). Therefore, our findings imply that the CAT-Q(-NL) may measure camouflaging intent while the discrepancy approach measures camouflaging efficacy, which has also been suggested by Cook and colleagues (2021).

Moreover, when using the discrepancy approach, we assume the AQ represents someone's internal autistic characteristics and the ADOS their external presentation. However, it is difficult to find an instrument that can represent someone's true internal autistic characteristics. That is, when using a self-report questionnaire such as the AQ the level of internal autistic characteristics may be biased by someone's own interpretation of their autism characteristics. Another approach of measuring internal autistic status is by measuring (social) cognition, without relying on self-report (Livingston, Colvert, et al., 2019). Future research could focus on the differences between these approaches and if there is a preference of one over the other. Also, someone's behavior on the ADOS might not be representative of someone's external autistic characteristics in daily life, because the ADOS is designed to "provoke" autistic behavior and participants might not feel the need to camouflage when participating in a study investigating autism. These are reasons why the validity of the discrepancy approach has been debated (Fombonne, 2020; Lai et al., 2021; Williams, 2022) but given its customary use in scientific research (e.g.(Corbett et al., 2021; Lai et al., 2016, 2018; Livingston, Colvert, et al., 2019; Parish-Morris et al., 2017; Ratto et al., 2018; Rynkiewicz et al., 2016; Schuck et al., 2019; Wood-Downie et al., 2020), it is important that we better understand if this approach indeed measures camouflaging behavior.

A strength of this study is that it is the first comparison of a self-report measure of camouflaging behavior with the discrepancy measure. Comparisons like this are essential to learn more about the camouflaging construct. Another strength is that we included a large heterogeneous sample for this study, by recruiting via mental health institutions and social media. In addition, as this study was part of a large research project investigating aging in autism, we included a more representative sample than if we would have advertised a study about camouflaging behavior. That is, studies investigating camouflaging behavior include on average twice as many females than males (Libsack et al., 2021), while the sex ratio in the present study is close to 1:1. The main limitation of our study is that convergent validity of the CAT-Q-NL is difficult to establish, because there is no "gold standard" instrument for camouflaging behavior. It would be insightful to compare the CAT-Q-NL to another recently developed camouflaging questionnaire such as the Compensation Checklist (Livingston et al., 2020) and to instruments that measure more general constructs related to camouflaging behavior, e.g., normative or social conformity. Developing instruments to investigate new constructs is a challenging process and can be seen as an iterative scientific cycle (Cronbach & Meehl, 1955). Therefore, we should continue to further disentangle the construct validity of instruments that can be used to measure camouflaging. Another limitation is that our results might not generalize to all autistic individuals as our participants were mainly diagnosed later in life, had no intellectual disability and -while we did not explicitly ask about ethnicity- we know that our sample was not culturally diverse. An interesting future research avenue is to determine how age of diagnosis and cultural background might impact camouflaging in autistic individuals.

4.1. Implications

Using the CAT-Q-NL we can gain insight in the level of camouflaging behavior of autistic individuals. However, the CAT-Q-NL is not suitable for comparing camouflaging behavior between autistic and non-autistic individuals. Further research should shed more light on the construct validity of the CAT-Q(-NL) by comparing results to other instruments that measure constructs similar to camouflaging behavior. As camouflaging behavior has been shown to be multilayered, we might even need additional measures to grasp all different layers of camouflaging behavior. Given the importance of camouflaging behavior for autistic individuals, the CAT-Q(-NL) is expected to be of value in clinical practice as well as in scientific research, despite the psychometric challenges. Using a questionnaire as the CAT-Q(-NL) ensures that an important topic as camouflaging is not overlooked by the clinician and autistic individuals themselves. This is especially important, as camouflaging is found to have negative consequences for autistic individuals and it can hinder proper

autism assessment and, subsequently receiving appropriate care.

CRediT authorship contribution statement

All authors contributed to the study design and pre-registration. Data was collected by W.J. van der Putten, T.A. Radhoe and C. Torenvliet. Analyses and writing of first draft of manuscript was performed by W.J. van der Putten. A.J.J. Mol and H.M. Geurts provided feedback on subsequent versions of the manuscript. All authors provided feedback on the final version and approved the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data that has been used is confidential.

Acknowledgements

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.rasd.2022.102072.

References

Agelink van Rentergem, J.A., Radhoe, T.A., Geurts, H.M. (2018). Dutch Translation of the Camouflaging Autistic Traits Questionnaire. [Unpublished].

Ai, W., Cunningham, W. A., & Lai, M.-C. (2022). Reconsidering autistic 'camouflaging' as transactional impression management. *Trends in Cognitive Sciences, 26*(8), 631–645. https://doi.org/10.1016/j.tics.2022.05.002

Bargiela, S., Steward, R., & Mandy, W. (2016). The experiences of late-diagnosed women with autism spectrum conditions: An investigation of the female autism phenotype. *Journal of Autism and Developmental Disorders, 46*(10), 3281–3294. https://doi.org/10.1007/s10803-016-2872-8

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

Bildt, A., Greaves-Lord, K., & De Jonge, M. (2013). .. *ADOS-2 Nederlandse bewerking.[ADOS-2 Dutch Adaptation]*. Amsterdam, Netherlands: Hogreve Uitgevers.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (Second ed.). Guilford Publications.

Cassidy, S. A., Gould, K., Townsend, E., Pelton, M., Robertson, A. E., & Rodgers, J. (2019). Is camouflaging autistic traits associated with suicidal thoughts and behaviours? Expanding the interpersonal psychological theory of suicide in an undergraduate student sample. *Journal of Autism and Developmental Disorders, 50* (10), 3638–3648. https://doi.org/10.1007/s10803-019-04323-3

Cook, J., Hull, L., Crane, L., & Mandy, W. (2021). Camouflaging in autism: A systematic review. *Clinical Psychology Review, 89*, Article 102080. https://doi.org/10.1016/j.cpr.2021.102080

Corbett, B. A., Schwartzman, J. M., Libsack, E. J., Muscatello, R. A., Lerner, M. D., Simmons, G. L., & White, S. W. (2021). Camouflaging in autism: Examining sex-based and compensatory models in social cognition and communication. *Autism Research, 14*(1), 127–142. https://doi.org/10.1002/aur.2440

Cronbach, L. J., & Meehl, P. E. (1955). *Construct Validity in Psychological Tests, 52*(4), 281–302. https://doi.org/10.1037/h0040957

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2009). *COTAN beoordelingssysteem voor de kwaliteit van tests (geheel herziene versie)*. NIP. http://www.psynip.nl/website/wat-doethet-nip/tests/beoordelingsprocedure/beoordelingsprocedure.

Fombonne, E. (2020). Camouflage and autism. *Journal of Child Psychology and Psychiatry, 61*(7), 735–738. https://doi.org/10.1111/jc

Geurts, H. M., Agelink van Rentergem, J. A., Radhoe, T., Torenvliet, C., van der Putten, W. J., & Groenman, A. P. (2021). Ageing and heterogeneity regarding autism spectrum conditions: A protocol paper of an accelerated longitudinal study. *BMJ Open, 11*(3), Article e040943. https://doi.org/10.1136/bmjopen-2020-040943

Groenman, A. P., van der Werf, S., & Geurts, H. M. (2022). Subjective cognition in adults with common psychiatric classifications; a systematic review. *Psychiatry Research, 308*, Article 114374. https://doi.org/10.1016/j.psychres.2021.114374

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191–205. https://doi.org/10.1177/1094428104263675

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the autism-spectrum quotient (AQ): A Study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders, 38*(8), 1555–1566. https://doi.org/10.1007/s10803-008-0538-x

Hull, L., Lai, M.-C., Baron-Cohen, S., Allison, C., Smith, P., Petrides, K. V., & Mandy, W. (2020). Gender differences in self-reported camouflaging in autistic and non-autistic adults. *Autism, 24*(2), 352–363. https://doi.org/10.1177/1362361319864804

Hull, L., Mandy, W., Lai, M.-C., Baron-Cohen, S., Allison, C., Smith, P., & Petrides, K. V. (2019). Development and validation of the camouflaging autistic traits questionnaire (CAT-Q). *Journal of Autism and Developmental Disorders, 49*(3), 819–833. https://doi.org/10.1007/s10803-018-3792-6

Hull, L., Petrides, K. V., Allison, C., Smith, P., Baron-Cohen, S., Lai, M.-C., & Mandy, W. (2017). "Putting on my best normal": Social camouflaging in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders, 47*(8), 2519–2534. https://doi.org/10.1007/s10803-017-3166-5

Hus, V., & Lord, C. (2014). The autism diagnostic observation schedule, module 4: Revised algorithm and standardized severity scores. *Journal of Autism and Developmental Disorders, 44*(8), 1996–2012. https://doi.org/10.1007/s10803-014-2080-3

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism, 20*(4), 442–462. https://doi.org/10.1177/1362361315588200

Kooij, J. J. S., Buitelaar, J. K., Oord, E. J. V. D., Furer, J. W., Rijnders, C. A. T., & Hodiamont, P. P. G. (2005). Internal and external validity of Attention-Deficit Hyperactivity Disorder in a population-based sample of adults. *Psychological Medicine, 35*(6), 817–827. https://doi.org/10.1017/S003329170400337X

Lai, M.-C., Hull, L., Mandy, W., Chakrabarti, B., Nordahl, C. W., Lombardo, M. V., … Livingston, L. A. (2021). Commentary: 'Camouflaging' in autistic people – reflection on Fombonne (2020). *Journal of Child Psychology and Psychiatry, 62*(8), 1037–1041. https://doi.org/10.1111/jc

Lai, M.-C., Lombardo, M. V., Chakrabarti, B., Ruigrok, A. N., Bullmore, E. T., Suckling, J., … Baron-Cohen, S. (2018). Neural self-representation in autistic women and association with 'compensatory camouflaging'. *Autism, 23*(5), 1210–1223. https://doi.org/10.1177/1362361318807159

Lai, M.-C., Lombardo, M. V., Ruigrok, A. N., Chakrabarti, B., Auyeung, B., Szatmari, P., … Consortium, M. A. (2016). Quantifying and exploring camouflaging in men and women with autism. *Autism, 21*(2), 133–141. https://doi.org/10.1177/1362361316671012

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (Fifth ed.). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Lawson, W. B. (2020). Adaptive morphing and coping with social threat in autism: An autistic perspective. *Journal of Intellectual Disability - Diagnosis and Treatment, 8* (3), 519–526. https://doi.org/10.6000/2292-2598.2020.08.03.29

Libsack, E. J., Keenan, E. G., Freden, C. E., Mirmina, J., Iskhakov, N., Krishnathasan, D., & Lerner, M. D. (2021). A systematic review of passing as non-autistic in autism spectrum disorder. *Clinical Child and Family Psychology Review, 24*(4), 783–812. https://doi.org/10.1007/s10567-021-00365-1

Livingston, L. A., Colvert, E., Bolton, P., & Happé, F. (2019). Good social skills despite poor theory of mind: Exploring compensation in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 60*(1), 102–110. https://doi.org/10.1111/jc

Livingston, L. A., Shah, P., & Happé, F. (2019). Compensatory strategies below the behavioural surface in autism: A qualitative study. *The Lancet Psychiatry, 6*(9), 766–777. https://doi.org/10.1016/S2215-0366(19)30224-X

Livingston, L. A., Shah, P., Milner, V., & Happé, F. (2020). Quantifying compensatory strategies in adults with and without diagnosed autism. *Molecular Autism, 11*(1), 15. https://doi.org/10.1186/s13229-019-0308-y

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205–223. https://doi.org/10.1023/A:1005592401947

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism Diagnostic Observation Schedule, second edition (ADOS-2) manual (Part I): Modules 1-4*. Torrance, CA: Western Psychological Services.

Lord, C., Rutter, M., & 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. https://doi.org/10.1007/BF02172145

Nuijten, S.E. & M.B.. (2018). statcheck: Extract Statistics from Articles and Recompute p Values (1.3.0). https://CRAN.R-project.org/package=statcheck.

Parish-Morris, J., Liberman, M. Y., Cieri, C., Herrington, J. D., Yerys, B. E., Bateman, L., … Schultz, R. T. (2017). Linguistic camouflage in girls with autism spectrum disorder. *Molecular Autism, 8*(1), 48. https://doi.org/10.1186/s13229-017-0164-6

Pearson, A., & Rose, K. (2021). A conceptual analysis of autistic masking: Understanding the narrative of stigma and the illusion of choice. *Autism in Adulthood, 3*(1), 52–60. https://doi.org/10.1089/aut.2020.0043

Putnick, D. L., & 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. https://doi.org/10.1016/j.dr.2016.06.004

Ratto, A. B., Kenworthy, L., Yerys, B. E., Bascom, J., Wieckowski, A. T., White, S. W., … Anthony, L. G. (2018). What about the girls? Sex-based differences in autistic traits and adaptive skills. *Journal of Autism and Developmental Disorders, 48*(5), 1698–1711. https://doi.org/10.1007/s10803-017-3413-9

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(1), 1–36. https://doi.org/10.18637/jss.v048.i02

RStudio Team. (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, *MA URL* http://www.rstudio.com/.

Rynkiewicz, A., Schuller, B., Marchi, E., Piana, S., Camurri, A., Lassalle, A., & Baron-Cohen, S. (2016). An investigation of the 'female camouflage effect' in autism using a computerized ADOS-2 and a test of sex/gender differences. *Molecular Autism, 7*(1), 10. https://doi.org/10.1186/s13229-016-0073-0

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29*(4), 347–363. https://doi.org/10.1177/0734282911406661

Schuck, R. K., Flores, R. E., & Fung, L. K. (2019). Brief report: Sex/gender differences in symptomology and camouflaging in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 49*(6), 2597–2604. https://doi.org/10.1007/s10803-019-03998-y

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*(3), 393–408. https://doi.org/10.1007/BF02294363

Tierney, S., Burns, J., & Kilbey, E. (2016). Looking behind the mask: Social coping strategies of girls on the autistic spectrum. *Research in Autism Spectrum Disorders, 23*, 73–83. https://doi.org/10.1016/j.rasd.2015.11.013

Wechsler, D. (2008). *Wechsler adult intelligence scale–Fourth Edition (WAIS–IV), 22* p. 1). San Antonio, TX: NCS Pearson.

Wechsler, D. (2012). WAIS-IV-NL: Wechsler adult intelligence scale–Nederlandstalige bewerking: Technische handleiding en Afname en scoringshandleiding. *BV APAaI, Editor* .

Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using iq test performance of minorities. *Educational Measurement: Issues and Practice, 29*(3), 39–47. https://doi.org/10.1111/j.1745-3992.2010.00182.x

Williams, Z. J. (2022). Commentary: The construct validity of 'camouflaging' in autism: Psychometric considerations and recommendations for future research - reflection on Lai et al. (2020). *Journal of Child Psychology and Psychiatry, 63*(1), 118–121. https://doi.org/10.1111/jc

Wood-Downie, H., Wong, B., Kovshoff, H., Mandy, W., Hull, L., & Hadwin, J. A. (2020). Sex/gender differences in camouflaging in children and adolescents with autism. *Journal of Autism and Developmental Disorders, 51*(4), 1353–1364. https://doi.org/10.1007/s10803-020-04615-z