Hechler *et al. Molecular Autism* (2025) 16:7
https://doi.org/10.1186/s13229-025-00643-7

Molecular Autism

RESEARCH

Open Access

"What does 'often' even mean?" Revising and validating the Comprehensive Autistic Trait Inventory in partnership with autistic people

Friederike Charlotte Hechler[1,2*], Outi Tuomainen[2], Nathan Weber[3], Frank Fahr[5], Bodie Karlek[7], Marie Maroske[4], Meike Misia[6] and Nathan Caruana[3*]

Abstract

Background In this study, we revised the comprehensive autistic trait inventory (CATI)—a self-report inventory of autistic traits, in collaboration with autistic people and provided preliminary evidence for its validity as a self-report measure of autistic traits in the general population. An established strength of the CATI is its ability to capture female autistic traits. Our project aimed to extend this further, to increase the inventory's accessibility, and to minimise stigma induced by deficit-based representations of autistic experience.

Methods Together with 22 individuals from the autism and autistic communities, we created the Revised Comprehensive Autistic Trait Inventory (CATI-R). Revisions included rewording items to increase clarity or reduce stigma and expanding items to capture diverse autistic experiences. We also present a series of guidelines for developing self-report inventories of subclinical neurodivergent traits. We validated the CATI-R within a large sample (n = 1439), comprising people with a self-reported autism diagnosis (n = 331), people who self-identified as autistic (n = 44), and non-autistic participants (n = 1046).

Results We successfully validated a revision of the CATI. A confirmatory factor analysis supported the six-subscale structure (two-factor bifactors model: Chi-squared = 2705.73, $p < .001$, RMSEA = .04, SRMR = .03, CFI = .95, TLI = .94). Spearman's rank correlations showed positive relationships between all subscales (all $rs > .56$, $ps < .001$). Convergent validity was demonstrated by significant correlations between the CATI-R and two contemporary inventories of autistic traits: the AQ (rho = .86, $p < .01$) and BAPQ (rho = .82, $p < .01$). Finally, a measurement invariance analysis indicated that total-scale scores can be compared across genders.

Limitations Our study presents only initial evidence for the validity of the CATI-R that should be enriched with further analyses and types of data, including a larger number of participants who do not identify as male or female.

Conclusions This project provides a revised trait inventory that resonates with actual autistic experience, along with guidelines for creating self-report measures that are sensitive, accessible, and non-stigmatising.

*Correspondence:
Friederike Charlotte Hechler
fhechler@uni-potsdam.de
Nathan Caruana
nathan.caruana@flinders.edu.au
Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Lay abstract

Autistic people know what it means to be autistic. So autistic people may be especially well-placed to determine how autistic traits and experiences should be measured and described. Until now, most autism questionnaires have been made by non-autistic people. In our project, autistic people—including those with and without academic research backgrounds—edited an autism questionnaire called the 'Comprehensive Autistic tTrait Inventory', or 'CATI' for short. This is a survey that requires people to read a list of statements and indicate how much each relates to their own experiences. It is used to measure the extent to which people in the general population (including those who are not autistic) have experiences associated with autism. In a large online study, we found that the edited measure consistently and accurately measured autistic traits. We also propose basic guidelines for developing measures that better capture autistic people's experiences by using questions that are respectful and follow language preferences of the community.

Keywords Autism spectrum, Autistic traits, Neurodiversity, Psychological testing, Psychometric validation, Participatory research, Gender disparity

Background

Self-report screening and assessment methods of autistic[1] traits play an important role in both research and practice. These tools are often brief, inexpensive, non-invasive, and low-burden screening tools [3] during the individual diagnosis process [4] or for characterising participants in research [5] and clinical practice [6]. At the same time, these inventories have a direct impact on society and the way autistic people are perceived because they are available online and used by interested members of the public (see [7] on the phenomenon of autistic self-diagnosis). As such, it is critical that these tools capture a wide range of autistic experiences and are easy to understand. This project aimed to improve these aspects for the Comprehensive Autistic Trait Inventory (CATI; [8]) in a collaborative project with individuals from the autism and autistic communities.

The most widely used inventory of autistic traits is the Autism-Spectrum Quotient (AQ; [9]), with more than 4506 citations on Scopus. Other commonly used and popular inventories are the Broad Autism Phenotype Questionnaire (BAPQ; [10]), and the adult version of the Social Responsiveness Scale (SRS; [11]), with 137 and 2920 citations on Scopus, respectively. Despite their widespread use, several problems have been identified with existing inventories of autistic traits. A key consideration is that these inventories are believed to be less valid and sensitive for characterising autistic traits in females[2] than in cis-gender males [13–15]. Belcher et al. [16], for instance, concluded that only two items of the AQ assess autistic traits similarly across female and male test takers.

Male biases in autistic trait inventories are, in part, a symptom of a broader historical bias in autism research, which has been based on studies involving mostly male autistic people (see [17–19]). This imbalance has led to a male-oriented conceptualisation of autism [20–22]. However, characteristics of female and male autistic people often differ across the two DSM-5 domains of (1) social communication/interaction and (2) restricted and repetitive behaviour (see [23] for a review). For instance, social communication behaviours in autistic females are more similar to those of non-autistic peers (e.g., [24–26]) with greater engagement and better experience of friendship quality [27]). Likewise, repetitive behaviours and interests are either less common and/or harder to detect in autistic females [28, 29]. Autistic females also engage in more 'masking' or 'camouflaging' behaviours to appear more neurotypical [13, 21, 30]). Furthermore, autistic females' non-verbal behaviour (e.g., facial expressions and eye contact; [13, 14]) and verbal behaviour (e.g., speaking rate; [31]) are more similar to those of their non-autistic peers. While autistic males tend to exhibit more externalising behaviours, such as hyperactivity, impulsivity, and conduct problems, autistic females more often face internalising challenges like anxiety, depression, and eating disorders [32, 33]. Such mental health difficulties may be associated with a greater social pressure on females to mask autistic traits [34], requiring cognitive effort that

[1] We would like to stress that, in contrast to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; [1]), we understand autism as a condition that includes strengths and difficulties rather than a disorder. Furthermore, we will follow preferences from parts of the autism community to use identity-first language in this paper (e.g. "autistic person") while recognising that some individuals prefer the use of person-first language (e.g. "person with autism"; [2]).

[2] Since previous research often did not clarify how they differentiated between 'sex' and 'gender' [12], we will use the term "female" broadly here referring to cisgender women, transgender, non-binary, and gender diverse people, as well as anyone who was socialised, or identifies as female.

often leads to meltdown, exhaustion, autistic burnout, and poor mental health outcomes in general [21, 35, 36]).

Autistic females and gender-diverse people show additional difficulties, such as cognitive or intellectual impairment [37], and they are more likely than males to be undiagnosed, misdiagnosed, or diagnosed late because of their differing experiences of autism [38–45]. These outcomes can be compounded when both researchers and clinicians continue to use measures that are based on stereotypical or male-oriented conceptualisations of autism [43, 46]. Growing recognition of gender differences in the autistic experience (e.g., [38–45]) has led to intensifying calls from the autism community for research on this topic (e.g., [47]).

The Comprehensive Autistic Trait Inventory

The Comprehensive Autistic Trait Inventory (CATI; [8]) is a new inventory of autistic traits designed to better represent female experience of autism than earlier inventories. It measures 'sub-threshold' autistic traits that resemble the features of autism that may be experienced by people who do not meet the diagnostic criteria for autism. The CATI comprises 42 items that assess six trait dimensions: 'Social Interactions', 'Communication Difficulty', 'Social Camouflage', 'Repetitive Behaviours', 'Cognitive Rigidity', and 'Sensory Sensitivity'. While some of these subscales (e.g. 'Communication Difficulty'), overlap with those in widely used inventories of autistic traits like the AQ [9] and the BAPQ [10], the Social Camouflage and Sensory Sensitivity subscales are unique to the CATI, making the questionnaire particularly sensitive for identifying autistic traits in females. English and colleagues[ ] presented evidence for the CATI's convergent validity when using both the total-scale and subscale scores. They also demonstrated high subscale reliability and evidence for consistent item interpretation in male and female respondents [8]. The original CATI validation study also demonstrated better internal reliability for total-scale scores compared to the AQ and the BAPQ and greater predictive ability for classifying autism (in a sample of self-identifying autistic people).

While the CATI represents an improvement from previous inventories—particularly in better-capturing the female autism phenotype—some autistic participants in our current research identified further aspects that could be improved. In particular, autistic people highlighted that the use of 'neurotypical language' in the CATI, and other autistic trait measures, means that some items may: (1) not fully capture the autistic experiences (e.g., [48, 49]); and (2) not be understood in the same way by autistic and non-autistic people (e.g., [50–52]); or (3) inadvertently stigmatise the autistic experience (e.g., [53, 54]).

To ensure that measures of autistic traits are accurate, accessible and mitigate stigma, autistic people themselves should be involved in the design of autism-related measures [55]. Autistic people often manifest strong interest in and knowledge about autism and have demonstrated important contributions through participatory research by increasing the validity and accuracy of research methods and dissemination [47, 54, 56–60] and by ensuring that research meets the needs and priorities of its stakeholders in a respectful way [61–63].

The current project

The current study was initiated by autistic people who participated in a recent study involving the administration of the CATI. Two key priorities emerged: (1) revising the language used in the CATI to improve accessibility and mitigate stigma and (2) increasing the measure's sensitivity to the experiences of female and gender-diverse people. In collaboration with autistic people, we aimed to co-design a revision of the CATI to address these priorities. In a second step, we evaluated the revised inventory's psychometric properties as a research tool to measure individual differences in sub-clinical autistic traits within the general population following the protocol of the original CATI [8].

Methods

Revision of the CATI

Between January and March 2023, we received feedback on the original CATI [8] and co-designed revised versions (see below) through an iterative revision process with 22 individuals contacted by the research team. These individuals included autistic collaborators (n = 13; some with autism-related work), a mother of an autistic child, and professionals working with autistic people as carers, therapists, and clinicians (n = 8). Four non-academic autistic people contributed significantly to the revision and became co-authors of this project. Our engagement process involved various methods tailored to their preferences and availability (e.g., in-person meetings, video calls, or email correspondence) and progressed through multiple stages of feedback and revision.

Based on this feedback, we made revisions to all but six CATI items and added three new items. Since our collaborators were located in Australia, Germany, Switzerland, and Italy, we translated the CATI into German, and Italian (including back-translations, see https://www.cati-autism.com/translations), to meet the language needs of all collaborators. In-depth descriptions and justifications of our revisions with a table that presents each item of the revised CATI (CATI-R) against the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; [1]) core criteria for autism can be found on the Open

Science Framework page of this project (https://osf.io/vt4d2/). The revisions comprised two key types of revision: rewording and expansion. Our rewording addressed the community research priorities of 'Respect' and 'Accessibility'.

Respect

- We adopted a neurodiversity framework (e.g., [54]) and replaced language that was deemed by autistic partners as exclusionary or unnecessarily negative. For example, we substituted "react poorly" with more neutral language ("have strong reactions"), and used "masking" instead of "camouflage", which was a preferred term among our collaborators. A reason for this preference was the fact that animals, in contrast to autistic people, camouflage naturally without spending great amounts of cognitive effort. Moreover, "camouflage", when used in a military context (with which our collaborators did not want to be associated), implies an intentional tactic or strategy. Masking, in contrast, does not always reflect an explicit strategy engaged by autistic people, and may be associated with a cognitive or affective cost—either during or after the masking experience (see e.g., [21, 35, 36]).
- We also aimed to mitigate the potential for items to propagate stigma or autism stereotypes such as the idea that autistic people have no desire or motivation for social interactions (see [64]). We also added autistic experiences not captured in the original CATI items, such as the experience of empathy overload [65].
- Whenever possible, we clarified whether behaviour referred to interactions with non-autistic people or autistic people. This distinction is important because some challenges arise only when individuals from these two groups interact (i.e., the 'Double Empathy Problem'; [51]). Autistic people might, for example, be misunderstood by non-autistic interlocutors but not necessarily by other autistic interlocutors (e.g., [66, 67]).

Accessibility

- We clarified language to make the CATI more accessible. We aimed to align the literal interpretation of each item with its intended meaning to mitigate the misinterpretation of items in an overly literal way [68]. Therefore, we avoided figurative language and metaphors, or quotation marks to indicate conditional word meanings, opting for explicit language instead. For instance, we replaced the figurative term "'unspoken rules' of social situations" with "norms" and removed the quotation marks.
- We removed abstract or imprecise terms, like "often", to reduce uncertainty and increase response reliability.
- When necessary, we broadened items by always giving at least two examples. This also ensured that the examples did not unintentionally narrow the perceived relevance of the item.
- We reformulated items that required an introspective analysis of one's behaviour (e.g., changing "I look for strategies to…" to "I make an effort to…"). We also added additional items that can be answered by individuals who are less aware of their behaviours or strategies. For example, we added an item that refers to the result of atypical behaviour during social interaction ("People who do not know me well tend to misunderstand me").

To capture diverse autistic experiences, we added new items as well as alternative examples within existing items. These changes better reflect the heterogeneity of the individual autistic phenotype (see e.g., [69, 70]). We merged items addressing the same aspect or experience to accommodate for the addition of new items while keeping the number of items per scale constant. For instance, the items "I find social interactions stressful" and "Social occasions are challenging for me" were merged into "I find social occasions challenging/tiring/stressful" (see the documentation of all revision steps on the Open Science Framework (https://osf.io/vt4d2/). This way, we extended, for example, the Masking subscale (referred to as "Social Camouflaging" in the original CATI) since we considered this aspect particularly important for capturing female autistic traits (see Background above). Moreover, previous inventories have often assumed that autistic traits are static or dispositional. However, there is now growing recognition that some autistic traits or experiences may change over time, to the extent that diagnostic criteria may no longer be met in some cases [71, 72]. Repetitive sensory and motor behaviours, for instance, seem to correlate negatively with age [73]. Potential reasons for changes in autistic traits include intervention, compensation strategies [74], masking, or increased autonomy and control over the environment. Therefore, we instructed people to consider the relevance of statements both in the present and across their lifetime.

Validation of the CATI-R

We conducted a validation study on the revised items using the same protocol reported by English and colleagues for validating the CATI [8]. Our protocol was pre-registered on the Open Science Framework (https://osf.io/vt4d2/).

Participants

As in the original CATI validation study [8], we recruited participants via Prolific (M_{age} = 37 years, SD = 13; 688 female, 659 male, 82 non-binary, 10 not reported). Participants had to be at least 18 years old, native English speakers, and reside in one of five English-speaking countries (UK, USA, Canada, Australia, New Zealand). We also required participants to have a minimum approval rate of 99 on Prolific (representing overall reliability of the participant across different online experiments). Prolific provides superior data quality compared to other online studies (e.g., [75, 76]). Consistent with the original validation protocol [8], we excluded participants from all analyses if they: (1) reported English as a non-primary language, (2) failed two or more of the four attention checks embedded within the survey, (3) completed the questionnaires too quickly (< five minutes in total).

This resulted in 42 exclusions (2.84%) and a final data set of 1439 participants (320 more than in the original validation study; see Table 1 for more detailed demographics). The remaining data set contained no missing data as the design of our survey required participants to respond to each item before proceeding to the next page.

We collected data in separate batches to give people from all included countries an equal chance to participate. Each batch was published at the same local time each day for one of the included countries (e.g., Batch 1: 6 pm for UK/London; Batch 2: 6 pm for USA/Los Angeles). Most participants resided in the UK (n = 828, 57.54%) and the USA (n = 409, 28.42%; see Table 1 for detailed information). We classified participants with respect to their native language, gender identity, self-reported diagnosis status, and autism identity based on responses to a demographic questionnaire (rather than their Prolific profile) to ensure that the information was correct at the time of data collection. Data collection proceeded in 14 separate batches. We started with a sample of 605 people without screening based on self-reported autism diagnosis. Next, we recruited 400 additional participants who were screened by Prolific as having a self-reported autism diagnosis. Moreover, we recruited a further 400 participants screened for not having a self-reported diagnosis (again, see Table 1). Finally, we recruited an additional set of individuals registered as non-binary on Prolific. Data from this group were collected for a separate exploratory analysis that comprehensively investigated gender differences on the CATI-R. This additional set comprised 38 participants (M_{age} = 33.31, SD = 10.98) but 12 of these participants did not indicate that they were non-binary on our questionnaire and were classified according to this response.

The final dataset included participants who self-reported: (1) identifying as autistic with a formal autism

Table 1 Exposure groups per country

Country of residenle	Number of participants					Age			
	n	Female	Male	Non-binary	Prefer-not-to-say	Minimum	Maximum	M	SD
No diagnosis									
Australia	62	19	41	2	0	20	64	33.93	11.58
Canada	104	53	48	2	1	20	70	35.71	12.32
New Zealand	13	5	8	0	0	24	72	41.23	15.27
United Kingdom	656	381	258	12	5	18	77	39.86	13.60
United States	271	102	141	26	2	18	73	35.17	11.58
NA	2	2	0	0	0	NA	NA	NA	NA
	1108	562	496	42	8	18	77	38.01	13.12
Diagnosis									
Australia	15	7	5	3	0	21	47	29.40	7.35
Canada	5	2	2	1	0	20	34	26.40	5.55
New Zealand	1	0	1	0	0	26	26	26.00	NA
United Kingdom	172	74	82	15	1	18	63	32.62	10.08
United States	138	43	73	21	1	19	59	31.73	9.39
	331	126	163	40	2	18	63	31.99	9.64
	1439	688	659	82	10	18	77	37.00	13.00

diagnosis (i.e., *diagnosis*: n = 331; 126 female, 163 male, 40 non-binary, 2 not reported); (2) identifying as autistic without a formal diagnosis (i.e., *self-identify*; n = 44; 18 female, 18 male, 7 non-binary, 1 not reported); (3) not identifying as autistic and without a formal diagnosis (i.e., *no diagnosis*: n = 1108; 562 female, 496 male, 42 non-binary, 8 not reported; see Table 1). Compared to the original CATI validation study, our sample comprised significantly more autistic people who reported having received a formal diagnosis (current: 23%; original: 0.8%). We did not combine the diagnosis and self-identify group in our analyses (see the analysis script on our Open Science Framework page: https://osf.io/vt4d2/). Table 2 shows the ethnic identity of participants.

Materials

The study was hosted on the PCIbex platform, an online tool for conducting behavioural research.

Demographic questionnaire Participants filled in a demographic questionnaire that collected information on age, sex, gender, primary language, self-reported autism diagnosis status, and autistic self-identity.

Autistic-traits questionnaires All participants filled in the CATI-R as well as the AQ and the BAPQ [8]. To ensure that participants read each item carefully, an attention-check item was placed at the midpoint of each questionnaire, directing participants to select a specific response.

The CATI-R In line with the original design [8], the CATI-R consists of 42 items divided into six subscales each comprising seven items: (1) Social Interactions (SOC); (2) Sensory Sensitivity (SEN); (3) Repetitive Behaviours (REP); (4) Communication Difficulty (COM); (5) Cognitive Rigidity (RIG); (6) and Masking (MAS; previously referred to as "Social Camouflaging"). Items are rated on a 5-point Likert scale, as in the original CATI [8]. The original order of items, which distributed items from the different subscales as well as negatively formulated items evenly, was maintained, except for the five items that were merged or substituted. The complete version of the CATI-R, along with its scoring key, can be found alongside our preregistration on the Open Science Framework (https://osf.io/vt4d2/).

The AQ We also administered the 50-item Autism Quotient (AQ), developed for assessing autistic traits in non-autistic people [9]. AQ items are allocated to five distinct trait dimensions designated as 'Social Skill', 'Attention Switching', 'Attention to Detail', 'Communication', and 'Imagination'. Participants responded using a 4-point Likert scale with approximately half of the items using a reversed scale. We followed the scoring method used in the original CATI validation study. Responses were allocated a score of 1–4 along the four-point score reflecting degree of endorsement. This scoring method has been shown to enhance item discriminability [77] compared to the dichotomous scoring method that is typically used for the AQ that only scores items as 'agreement' (1) or 'disagreement' (0).

The BAPQ Finally, we administered the Broad Autism Phenotype Questionnaire (BAPQ), which is a 36-item self-report questionnaire. It was originally created to assess autistic traits in close relatives of autistic people [10], but it is now widely used to measure variation of autistic traits in the general population. The items are evenly distributed among three subscales: 'Pragmatic Language', 'Aloof Personality', and 'Rigid Personality'. For each task, participants responded on a 6-point Likert scale, with approximately half of the items being reverse-coded. Responses were scored from 1 to 6, with higher scores indicating greater endorsement of autistic traits.

Procedure

The study was approved by the University of Potsdam's Human Ethics Committee (protocol number: 84/2021), which is compliant with the Declaration of Helsinki and its later amendments. After providing consent, participants completed the demographic questionnaire and, subsequently, the three questionnaires on autistic traits in a randomised order. Each questionnaire was introduced with a separate task instruction. Participants were reimbursed with £ 5.25 for an average duration of 14 min.

Table 2 Ethnic identities across participant groups (percentage of the total sample)

	Asian	Black	Mixed	Other	White	No response
No diagnosis	5.98	4.17	3.75	1.39	60.95	0.76
Diagnosis	0.42	1.04	1.53	0.76	19.11	0.14
Total	6.40	5.21	5.28	2.15	80.06	0.90

Analysis

We implemented the same analyses undertaken in Study 2 of the original CATI validation process [8], as well as some further exploratory tests. We will report here only the subset of these analyses relevant to validate the CATI-R as a measure of individual differences in sub-clinical autistic traits within the general population. We have also conducted preliminary analyses exploring the potential predictive ability of the CATI-R as a screening tool, which are reported as a supplementary material within the project's Open Science Framework project page (https://osf.io/vt4d2/).

All statistical analyses were performed using R Statistical Software (v4.3.0; R Core Team 2023). Necessary assumptions including normality and homogeneity of variance, were assessed (see the Open Science Framework page of this project: https://osf.io/vt4d2/). Given the large sample size, violations of normality were deemed to have minimal impact on the results [78]. When assumptions were in doubt, appropriate steps were taken, including the use of non-parametric tests and data transformations.

Confirmatory factor analyses

To be an adequate inventory of autistic traits, the CATI-R items should cluster into groups related through latent variables or factors that conform with theory of the underlying 'autism' construct. To assess the fit between observed data and the theoretically grounded model of autism, we tested correlations between responses of the CATI-R using Confirmatory Factor Analysis (CFA). We conducted CFA, using the factor structures assumed to be underlying the original CATI [8]:

- A single, general factor for 'autistic traits'.
- Two dimensions informed by the core diagnostic criteria for autism (social communication/interaction domain and restricted and repetitive behaviour domain) labelled as 'social traits' (based on the subscales SOC, COM, and MAS) and 'non-social traits' (based on the subscales SEN, REP, and RIG).
- Six dimensions currently associated with autism, featured in the six questionnaire subscales of SOC, SEN, REP, COM, RIG, and MAS.

We entered these factors in the same statistical models tested in the original CATI validation study [8]:

(a) *One-factor model* A model comprising a single general factor.
(b) *Correlated Two-factor model* A model comprising a social (SOC, COM, MAS) and a non-social (REP, RIG, SEN) factor.
(c) *Single hierarchical factor model* A model comprising six factors (SOC, COM, MAS, REP, RIG, SEN) as part of a higher-order factor.
(d) *Correlated social and non-social hierarchical factors model* A model separately grouping the social (SOC, COM, MAS) and non-social (REP, RIG, SEN) factors at the second-order level.
(e) *Six-factor bifactor model* A model in which six factors exist in tandem with a bifactor.
(f) *Two-factor bifactors model* A model in which the social and non-social factors exist in tandem with two bifactors.
(g) *Correlated six-factor model* A model comprising six correlated factors.

First-order factors were derived from the covariation among the observed variables. Second-order factors accounted for the covariation among the multiple first-order factors. The CFA was conducted using the `lavaan` R package (v0.6.17; [79]) with polychoric correlations and weighted least squares estimation. Goodness-of-fit for the models was evaluated against robust close-fit indices, including the root mean square error of approximation (RMSEA) and standardised root mean square residual (SRMR). RMSEA and SRMR values below .08 were considered indicative of fair fit, and values below .06 were considered indicative of good fit. In addition, the comparative fit index (CFI) and Tucker–Lewis index (TLI), which range from 0 to 1, were also used as measures of model fit. CFI and TLI values above .90 were considered indicative of fair fit and values above .95 were considered indicative of good fit [80]. We also reported for model evaluation the Akaike information criterion (AIC). If the difference between models was between 2 and 7, we considered the model with the lower AIC as moderately better and if the difference was greater than 10 we considered it to be significantly better [81]. We expected that, as in the validation study of the original CATI [8], fit indices would support the six subscales structured within a bifactor model with second-order factors separately encompassing items loading on a social and non-social subscale.

Besides assessing overall model fit, we considered the strength of the item loadings on the intended factors. Greater factor loadings indicate that a variable is correlated with the respective factor [82]. The following thresholds for individual item loadings have been recommended [83]: .71 (50% overlapping variance) = excellent, .63 (40% overlapping variance) = very good, .55 (30% overlapping variance) = good, .45 (20% overlapping

variance) = fair, and .32 (10% overlapping variance) = poor.

Correlations between CATI-R subscales

To address potential concerns arising from non-normality, we tested for potential relationships between CATI-R subscale scores using Spearman's rank correlation tests. We expected a strong correlation between all subscales, similar to the validation study of the original CATI [8].

Internal consistency

To measure the internal reliability of the CATI-R subscale scores, we first calculated Cronbach's α, setting a threshold of .80 or higher as a value indicating adequate internal consistency. In addition to Cronbach's α, we also calculated the average inter-item correlation. We assumed an average inter-item correlation in the range of .15 to .50 to be desirable [84].

For the total scale, we computed McDonald's omega hierarchal, an alternative index based on inter-subscale correlations rather than inter-item correlations [85], and Cronbach's α stratified across subscales [86]. This index mitigated the potential overestimation of total-scale internal consistency for the multi-dimensional CATI-R when using Cronbach's α [87]. We expected internal consistency across the scale and subscales to be as good as or better than for the original validation where Cronbach's α ranged from .81 to .94, McDonald's omega hierarchal was .81, and Cronbach's α stratified across subscales was .95.

Convergent validity evidence

To provide evidence for the CATI-R's convergent validity, we calculated the correlation between the CATI-R total-scale score and total scores of the AQ and the BAPQ using Spearman's rank correlation. We expected to establish convergent validity, evidenced by high correlations between CATI-R scores and AQ and BAPQ scores, similar to those observed in the CATI validation study [8].

Examination of gender differences

One of our aims during the revision process was to make the inventory more sensitive for people not fitting the traditional male-biased conceptualisation of autism, including female and non-binary people. Since our focus was on behavioural traits and psychosocial factors of gender, rather than physiological factors, we investigated gender differences rather than biological sex differences. This also allowed us to consider data from non-binary participants. Previous studies evaluating inventories of autistic traits only compared biologically female and male individuals (e.g., [8, 16]), whereas autistic people are more likely than non-autistic people to be gender diverse [88, 89]. Therefore, we conducted analyses on two sample subsets. The first only included male and female participants (i.e., male vs. female) to closely follow the original CATI validation study, and the second included non-binary participants (i.e., male vs. female vs. non-binary).

Table 3 AIC values for all models tested in the CFA (smaller values indicate better model fit)

Model	AIC
Two-factor bifactors model	174,829.2
Six-factor bifactor model	174,854.3
Correlated six-factor model	176,363.2
Correlated social and non-social hierarchical factors model	176,422.4
Single hierarchical factor model	176,451.1
Correlated two-factor model	181,365.4
One-factor model	182,914.1

Total scale and subscale scores To compare mean differences in total-scale as well as subscale scores across genders, we conducted separate Mann–Whitney U tests for the female versus male data using the `stats` R package (v4.3.0). For the data including female, male, and non-binary data, we conducted Kruskal-Wallis tests with post-hoc pairwise comparisons using the Wilcoxon rank-sum test using the `rstatix` R package (v0.7.2; [90]).[3] To investigate the effect of gender, subscale (SOC, COM, MAS, RIG, REP, SEN), and their interaction on individual responses, we fitted linear-mixed models for both sample subsets (female vs. male; female vs. male vs. non-binary). The models were fitted using restricted maximum likelihood estimation. All *p*-values were estimated using the `lmerTest` R package (v3.1.3; [91]), with a significance criterion of $\alpha < 0.05$. Consistent with original validation findings, we expected to find differences between male and female participants in the form of higher scores on the COM and REP subscales for males compared to females, and higher scores on the SEN subscale for females compared to males [8].

Measurement invariance analysis (female versus male) For meaningful group comparisons between female, male, and non-binary test takers, measurement invariance is required [92]. That is, items should function similarly across groups. Two indices for comparing classification agreement among groups are sensitivity and specificity. Using the model with the closest fit in our CFA, we tested for measurement invariance in different consecutive steps. In each, we fitted a more restricted model:

[3] The originally preregistered analysis of gender differences, which was based on the analysis of the original CATI validation, were revised following a suggestion from a reviewer.

Table 4 Robust fit indices for the seven models tested in the CFA

Model	Chi square	p	RMSEA	SRMR	CFI	TLI	Lower bound RMSEA CI	BIC
One-factor model	10,879.55	< .001	.09	.07	.72	.71	.09	183,356.9
Correlated two-factor model	9328.86	< .001	.09	.07	.76	.75	.09	181,813.5
Single hierarchical factor model	4404.53	< .001	.06	.06	.90	.89	.06	176,925.5
Correlated social and non-social hierarchical factors model	4373.90	< .001	.06	.06	.90	.90	.06	176,902.2
Six-factor bifactor model	2678.66	< .001	.04	.03	.95	.94	.04	175,577.8
Two-factor bifactors model	2705.73	< .001	.04	.03	.95	.94	.04	175,597.6
Correlated six-factor model	4298.70	< .001	.05	.06	.90	.90	.06	176,885.1

(1) *Configural invariance* refers to the equivalence of the basic structure of the measurement model (i.e., the number of factors and the items associated with each factor) across groups.
(2) *Metric/weak factorial invariance* tests whether the relationship between each item and the underlying construct is similar across groups, as indicated by equivalent factor loadings across those groups.
(3) *Scalar/strong factorial invariance* refers to the equivalence of item intercepts or thresholds (i.e., whether mean differences in the construct capture all mean differences in the shared variance of the items).
(4) *Residual/strict invariance* compares the sum of specific variance (variance of the item that is not shared with the factor) and error variance (measurement error) across groups.
(5) *Mean invariance* constrains the factor loadings, intercepts, and means to be equal across groups.

For each step, we report χ^2, CFI, and RMSEA difference tests, comparing the current model with the previous one. ΔCFI >.01 and ΔRMSEA >.015 were considered to be indicative of a violation of the invariance assumption [93, 94]. We expected, as in the original CATI validation study [8], to find evidence for measurement invariance.

Exploratory analyses

Although neither the CATI nor the CATI-R was originally developed as a diagnostic tool, we conducted preliminary investigations into the ability of the CATI-R to differentiate between autistic (diagnosis) and non-autistic people (no diagnosis). To this end, we performed a series of analyses, including linear-mixed models, logistic regression, hierarchical logistic regression, and Youden's Index. The detailed results, available on the project's Open Science Framework page (https://osf.io/vt4d2/), suggest that the CATI-R shows potential as a screening tool with distinct cut-off scores for female, male, and non-binary people. However, further research is required to validate its use as a screening tool, particularly through the administration of a gold-standard autism measure, such as the ADOS-2 [95], in conjunction with the CATI-R. The current study was unable to verify diagnoses due to the sample size constraints required for the primary analyses.

Results

Confirmatory factor analysis

To assess whether the CATI-R was consistent with current models of autism, we conducted CFA that identified the same two models as the best fit to the data as the original CATI validation study [8]. As reported in Table 3, the Two-Factor Bifactors Model had the smallest AIC value (AIC = 174,829), indicating significantly better model fit than the model with the second lowest AIC value, the Six-Factor Bifactor Model (AIC = 174,854).

A multivariate normality test indicated that, while the skewness assumption was held, the kurtosis assumption was violated (Mardia's test of kurtosis, $p < .01$), which was not surprising, given that the original CATI is a hierarchically multidimensional scale (i.e., unidimensional at the higher-order level, with correlated subfactors at the lower-order level). Although CFA is known to be relatively robust to moderate violations of kurtosis, particularly when the sample size is large (i.e., > 200; [78]), we addressed the non-normality of the data using Robust Maximum Likelihood estimation, which adjusts standard errors and chi-square statistics to account for non-normality. The Two-Factor Bifactors Model exhibited good model fit ($\chi^2 = 2678.66$, $df = 761$, NFI = .93, CFI = .95, TLI = .94, SRMR = .031, RMSEA = .042 with 90% CI 0.04–0.044; see Table 4),[4] with CFI (.95), SRMR (.031) and RMSEA (.042) values above the recommended threshold for good fit [96]. The TLI difference between

[4] The same model in the original CATI validation had the following fit indices: $\chi^2 = 3172.37$, $p < .001$, *CFI* = .955, and *TLI* = .950, *SRMR* = .054, *RMSEA* = .053 [8].

Table 5 Range of factor loadings across subscales for the Two-Factor Bifactors Model

Scale	Minimum	Maximum
Social traits		
SOC	.43	.61
COM	.07	.54
MAS	.20	.55
Non-social traits		
RIG	.18	.57
REP	.28	.54
SEN	.26	.64

SOC Social Interactions, *COM* Communication Difficulty, *MAS* Masking, *RIG* Cognitive Rigidity, *REP* Repetitive Behaviours, *SEN* Sensory Sensitivity

the two best-fitting models was − .001 and, thus, below the threshold of .01 for practical improvement [80].

We further evaluated each item by assessing whether its factor loading was sufficiently large to predict the item response based on the intended underlying factor in the Two-Factor Bifactors Model [83]. At least four items in the social traits and non-social traits factors had loadings of .60 or higher [97]. Specifically, we found that λ values on the social traits factor ranged from .08 to .76 with 39 of 42 items showing at least 'fair' loadings. The λ values on the non-social traits factor ranged from .42 to .77 with 40 of 42 items showing at least 'good' loadings. The ranges of λ values on each subscale factor can be found in Table 5. Items 11, 20, 21, 22, 24, 25, 26, 31, 32, 34, 39, 43, 44, 47 only achieved 'poor' loadings on their respective subscale factor ($\lambda < .32$; [83]). However, each of these items had good to excellent loadings on the respective social traits and non-social traits factor, indicating that these items loaded higher on one of the other subscales (see the Open Science Framework page of this project: https://osf.io/vt4d2/ for the individual loadings).

Correlations between CATI-R subscales

As expected from the original CATI validation study [8], all subscales were significantly positively correlated (see Table 6). Following the Bonferroni correction for multiple comparisons ($\alpha = .05/28 = .002$), these correlations remained significant (all adjusted ps < .001). We observed less variation than the original CATI validation. The magnitude of these correlations ranged only between $r = .56$ and $r = .68$, whereas the original validation study observed a range from $r = .26$ and $r = .55$ [8]. Thus, for the CATI-R, the associations between trait dimensions were more uniform than those of the CATI subscales. Again, as in the original validation study [8], we also found a significant correlation between the social and non-social bifactors ($r = .79$) and this correlation was higher than the one in the original CATI validation study ($r = .58$).

Table 6 Spearman's rank correlations between CATI-R subscales

Subscale 1	Subscale 2	Spearman's rank correlation coefficient	Statistic	*p* value
SOC	COM	.67	164,505,476	< .001
SOC	MAS	.57	214,774,542	< .001
SOC	RIG	.57	215,932,001	< .001
SOC	REP	.56	216,080,216	< .001
SOC	SEN	.57	213,875,696	< .001
SOC	Social traits	.88	60,673,401	< .001
SOC	Non-social traits	.64	176,596,871	< .001
COM	MAS	.60	199,274,674	< .001
COM	RIG	.60	200,331,788	< .001
COM	REP	.62	189,389,502	< .001
COM	SEN	.59	203,527,374	< .001
COM	Social traits	.88	61,526,609	< .001
COM	Non-social traits	.68	157,055,015	< .001
MAS	RIG	.63	185,097,878	< .001
MAS	REP	.67	164,497,853	< .001
MAS	SEN	.59	202,336,859	< .001
MAS	Social traits	.82	91,227,903	< .001
MAS	Non-social traits	.72	140,766,957	< .001
RIG	REP	.67	165,921,552	< .001
RIG	SEN	.61	191,486,898	< .001
RIG	Social traits	.69	155,179,712	< .001
RIG	Non-social traits	.85	75,023,613	< .001
REP	SEN	.68	157,855,798	< .001
REP	Social traits	.71	142,429,398	< .001
REP	Non-social traits	.90	49,955,323	< .001
SEN	Social traits	.68	160,251,930	< .001
SEN	Non-social traits	.88	61,112,245	< .001
Social traits	Non-social traits	.79	105,113,446	< .001

SOC Social Interactions, *COM* Communication difficulty, *MAS* Masking, *RIG* Cognitive Rigidity, *REP* Repetitive Behaviours, *SEN* Sensory Sensitivity

Internal consistency

We tested how closely related the items of each subscale were by calculating Cronbach's α. All CATI-R subscales exhibited high internal consistency (i.e., α >.84; see Table 7). These values were greater than those observed for the AQ and the BAPQ in our sample. All Cronbach's α values were comparable to those observed in the original CATI validation study [8] and, for four subscales as well as for the social and non-social bifactors, the CATI-R Cronbach's α values exceeded those observed for the original CATI. We further investigated the internal consistency of each measure and subscale for the CATI-R, AQ and BABQ separately for autistic and non-autistic participants. This revealed comparably high

Table 7 Cronbach's alpha across questionnaires

Subscale	Cronbach's alpha
CATI-R	
Social Interactions	.89
Communication Difficulty	.87
Repetitive Behaviours	.84
Cognitive Rigidity	.89
Sensory Sensitivity	.90
Masking	.89
Social bifactor	.93
Non-social bifactor	.94
AQ	
Social Skill	.88
Attention Switching	.83
Attention to Detail	.76
Communication	.85
Imagination Items	.76
BAPQ	
Aloof	.85
Pragmatic Language	.83
Rigid	.89

internal consistency of the CATI-R in both autistic and non-autistic samples (see the analysis script on the Open Science Framework page of this project: https://osf.io/vt4d2/).

When all 42 items were evaluated in a single factor, the internal consistency remained high, as assessed by McDonald's omega hierarchical and stratified Cronbach's α (see Table 8). These values were, again, higher than those of the AQ and the BAPQ in our sample and slightly higher than those of the CATI in the original validation study [8].

We further examined the extent to which scores on one item were related to scores on all other items to rule out item redundancy. As Table 9 shows, the average inter-item correlation value for all three questionnaires was in the desirable range of .15 to .50 [84].

Table 8 McDonald's omega hierarchical and stratified Cronbach's alpha across questionnaires

Total scale	McDonald's omega hierarchical	Stratified Cronbach's alpha
CATI-R	.82	.96
AQ	.67	.93
BAPQ	.69	.92

Table 9 Average inter-item correlation for the CATI-R, the AQ, and the BAPQ

Questionnaire	Average inter-item correlation
CATI-R	.38
AQ	.22
BAPQ	.26

Convergent validity evidence

We further tested how strongly CATI-R scores were correlated with AQ and BAPQ scores, two questionnaires assumed to measure the same constructs. As for the original CATI [8], the CATI-R correlated at the total-scale score level with both the AQ (ρ = .86, S = 71,117,196, p <. 01) and the BAPQ (ρ = .82, S = 86,936,630, p < .01). These correlations were similarly high as those reported in the original CATI validation study [8], providing evidence for convergent validity.

Examination of gender differences

Total scale and subscale scores by gender

Male versus female Our revision aimed to further extend the gender representativeness of the CATI. Thus, we evaluated how CATI-R results differed across genders. On the total scale, male participants scored significantly higher (M_{Male} = 129.99) than female participants (M_{Female} = 128.85; W = 67,621,853, p = .003), although the effect size was small, r = − .08.

To investigate whether gender (female, male), subscale (SOC, COM, MAS, RIG, REP, SEN), and their interaction had an effect on individual responses, we ran a linear-mixed model on transformed data, with a random intercept for participants to account for repeated measurements. The main effect of gender was non-significant (β = − 0.03, SE = 0.04, p = 0.53), indicating no difference in individual responses between female and male participants. However, the effect of the individual subscales was significant, and subscale interacted with gender in that females endorsed items on the SEN subscale more than males (β = 0.2, SE = 0.05, p < .01). However, the associated effect size was small (R^2 < = .001).

As Table 10 shows, male and female participants differed in some of the subscale scores. Separate Mann–Whitney U tests for each subscale score can be found in Table 11. Male participants scored significantly higher than female participants on the COM subscale, with a medium effect size (r = −.10), and female participants scored significantly higher than males on the SEN subscale, with a large effect size (r = −.14).

Table 10 Mean score per gender group across subscales

	COM	MAS	REP	RIG	SEN	SOC
Male	18.72	22.70	21.29	24.03	20.02	24.09
Female	17.49	22.50	20.74	23.39	22.24	23.57
Non-binary	23.34	26.17	27.06	26.88	27.94	27.85
Prefer-not-to-say	19.60	24.30	22.50	22.70	26.10	26.40
Total	18.40	22.81	21.36	23.87	21.58	24.07

COM Communication Difficulty, *MAS* Masking, *REP* Repetitive Behaviours, *RIG* Cognitive Rigidity, *SEN* Sensory Sensitivity, *SOC* Social Interactions

Table 11 Mann–Whitney U test: subscale score by gender (male, female)

Comparison	Statistics	Effect size
SOC score		
Female versus male	W = 234,915, $p = .25$	r = − .03
COM score		
Female versus male	W = 252,419, $p < .001$	r = − .10
MAS score		
Female versus male	W = 230,966, $p = .55$	r = − .02
RIG score		
Female versus male	W = 237,081, $p = .15$	r = − .04
REP score		
Female versus male	W = 235,124, $p = .24$	r = − .03
SEN score		
Female versus male	W = 190,330, $p < .001$	r = − .14

SOC Social Interactions, *COM* Communication Difficulty, *MAS* Masking, *RIG* Cognitive Rigidity, *REP* Repetitive Behaviours, *SEN* Sensory Sensitivity

Male versus female versus non-binary Non-binary participants had the highest mean score of all three gender groups ($M_{Male} = 155.81$). A Kruskal-Wallis test was conducted to compare total scores across the three gender groups. This revealed a statistically significant difference in scores between the groups ($\chi^2(2) = 676.31$, $p < .001$), with a large effect size ($\eta^2 = .47$). Post-hoc pairwise comparisons using the Wilcoxon rank-sum test indicated significant differences between female and non-binary ($p < .001$) and between male and non-binary participants ($p < .001$), with non-binary participants scoring significantly higher than both female and male. The effect size for these two comparisons was large ($r < -.66$), likely due to the small number of non-binary participants in our sample. Male and female participants differed as well ($p = .01$). However, the effect size was small ($r < -.07$).

As for the male versus female comparison, we investigated whether gender (female, male, non-binary), subscale (SOC, COM, MAS, RIG, REP, SEN), and their interaction had an effect on individual responses. We ran a linear-mixed model on transformed data, with a random intercept for participants to account for repeated measurements. Non-binary participants had significantly higher scores than males ($\beta = -.3$, $SE = .09$, $p < .01$), with a small effect size ($R^2 = .001$) and females ($\beta = -.41$, $SE = .09$, $p < .01$), with a small effect size ($R^2 = .001$). The effect of the individual subscales was significant and subscale interacted with gender in that non-binary participants endorsed items on the SEN subscale more than males ($\beta = -.38$, $SE = .1$, $p < .01$). The effect size was small ($R^2 < .001$).

Separate Kruskal–Wallis tests with pairwise comparisons using the Wilcoxon rank-sum test for each subscale score can be found in Table 12. Non-binary participants scored significantly higher than both females and males on all subscales. However, effect sizes were mostly small ($r < .03$).

Measurement invariance analysis (female vs. male)

We entered the model that showed the closest fit to our data in the CFA (two-factor bifactors model) in a measurement invariance analysis with respect to gender. Following the original CATI validation process [8], we only report here the analysis on female and male data, using data from the 1347 participants. The autistic group consisted of 289 participants, with 126 identifying as female and the non-autistic group consisted of 1058 participants, with 562 identifying as female. This analysis was then repeated, including also non-binary participants (for results, see the Open Science Framework page of this project: https://osf.io/vt4d2/).

We found evidence for configural invariance since the fit indices from the configural model (see Table 13) were aligned with the guidelines for good model fit [96]. Thus, the latent factors of autistic traits had the same pattern of free and fixed loadings across the two genders. When we constrained the loadings across groups and compared this metric model to the previous configural model (a test for metric invariance), the scaled χ^2 test was significant ($p < .001$) but the ΔCFI was < .01 and the ΔRMSEA was < .015 [93, 94]. Since χ^2 is sensitive to sample size, we assumed that the CATI-R still exhibited metric invariance. In other words, all items contributed to the construct to a similar degree across female and

Table 12 Kruskal–Wallis and pairwise Wilcoxon rank-sum tests: subscale score by gender (female, male, non-binary

Comparison	Statistics	Effect size
SOC score		
	Chi-squared (2) = 27.73, p < .001	Eta squared = .02
Female versus non-binary	U = 18,448, p (adjusted) < .001	r = − .13
Male versus non-binary	U = 18,126, p (adjusted) < .001	r = − .12
Female versus male	U = 234,915, p (adjusted) < .001	r = − .01
COM score		
	Chi-squared (2) = 48.43, p < .001	Eta squared = .03
Female versus non-binary	U = 16,185, p (adjusted) < .001	r = − .16
Male versus non-binary	U = 17,285.5, p (adjusted) < .001	r = − .14
Female versus male	U = 252,419, p (adjusted) < .001	r = − .09
MAS score		
	Chi-squared (2) = 27.34, p < .001	Eta squared = .02
Female versus non-binary	U = 18,582.5, p (adjusted) < .001	r = − .13
Male versus non-binary	U = 17,806, p (adjusted) < .001	r = − .13
Female versus male	U = 230,966, p (adjusted) < .001	r < .001
RIG score		
	Chi-Squared (2) = 21.37, p < .001	Eta squared = .01
Female versus non-binary	U = 19,572, p (adjusted) < .001	r = − .11
Male versus non-binary	U = 19,731, p (adjusted) < .001	r = − .10
Female versus male	U = 237,081.5, p (adjusted) < .001	r = − .02
REP score		
	Chi-squared (2) = 46.86, p < .001	Eta squared = .03
Female versus non-binary	U = 15,560, p (adjusted) < .001	r = − .17
Male versus non-binary	U = 15,190.5, p (adjusted) < .001	r = − .17
Female versus male	U = 235,124, p (adjusted) < .001	r = − .01
SEN score		
	Chi-squared (2) = 82.76, p < .001	Eta squared = .06
Female versus non-binary	U = 16,350, p (adjusted) < .001	r = − .16
Male versus non-binary	U = 11,528, p (adjusted) < .001	r = − .22
Female versus male	U = 190,330, p (adjusted) < .001	r = − .13

SOC Social Interactions, *COM* Communication Difficulty, *MAS* Masking, *RIG* Cognitive Rigidity, *REP* Repetitive Behaviours, *SEN* Sensory Sensitivity

Table 13 Multi-group factorial analysis assessing measurement invariance of the CATI-R as a function of participant gender (female vs. male)

Model	Chi^2	df	CFI	RMSEA	(90% CI)	Delta Chi2	Delta df	Delta CFI	Delta RMSEA	Decision
Configural invariance	2404.61	1524	1.00	.03	(.03–.03)	–	–	–	–	Accept
Metric (weak) invariance	2800.81	1601	1.00	.03	(.03–.04)	− 396.20	− 77.00	.001	− .004	Accept
Scalar (strong) invariance	2906.06	1716	1.00	.03	(.03–.03)	− 105.25	− 115.00	< .001	.001	Accept
Residual (strict) invariance	2906.06	1716	1.00	.03	(.03–.03)	< .001	< .001	< .001	< .001	Accept
Mean invariance	3367.76	1723	1.00	.04	(.04–.04)	− 461.70	− 7.00	.001	− .01	Accept

male participants. Next, we tested for scalar invariance, which examines whether the item intercepts are equivalent across groups. We found no evidence for reduced model fit in the scalar invariance model compared to the metric invariance model. Then, we established residual invariance, or equivalence of item residuals by further constraining the factor loadings, intercepts, and residual variances to be equal across groups. Finally, we also constrained means to be equal across groups. Both ΔCFI and ΔRMSEA were, again, below the thresholds implying that

the invariance assumption still held across all stages of the analysis.

Discussion

Autistic trait inventories play an important role in research and practice, as well as the public portrayal of autism. However, they are not always in line with current academic research, can be biased or stigmatising, and may not be easily accessible for autistic people, or those who experience prominent autistic traits. In this participatory project with autistic collaborators, we revised the CATI [8] and provided preliminary evidence of its strong psychometric properties as a measure of individual differences in autistic traits.

The collaborative revision of the CATI-R as a positive experience

Prompted by the autistic community members, we revised the CATI using a participatory approach. Collaborators judged the revised inventory as easier to interpret, less stigmatising, and more accurate in capturing autistic traits than the original CATI. In particular, female and non-binary contributors reported that their involvement in the project helped them advocate for themselves. One non-binary autistic collaborator noted:

> *I just read your revisions to the questionnaire, and I'm really impressed. You use considerate and understandable vocabulary, avoiding confusion or too many interpretation challenges (e.g., through concrete examples or precise explanations of what 'often' or 'over-sensitive' could mean). I'm confident that future autistic participants will be able to engage well with the test, and many questions have been clarified.*

The current study took the first critical step to involve autistic people in the development and evaluation of trait measures that directly relate to their lived experience. This has resulted in a measure that more accurately and sensitively captures autistic traits and experiences, can be appropriately understood by both autistic and non-autistic participants, represents the experiences of people with different gender identities, and mitigates the stigmatisation of autism. In doing so, this study has extended the many benefits of participatory autism research (e.g., [47, 57–59, 62, 98]), to the science of psychometric tools used in autism research and practice.

Empirical support for the validity of the CATI-R as a measure of individual differences in autistic traits in the general population

Our validation analyses largely replicated the results from the original CATI validation study [8]. First, CFA supported the six-subscale structure of the CATI-R and suggests that it is in line with the current diagnostic criteria for autism, which are grouped into social and non-social traits [1]. For the best-fitting model, several items showed poor loadings on their respective subscale. Some of these items might load onto multiple factors, given the correlations we observed, or load better onto another factor. In fact, these items loaded highly on the respective social or non-social traits factor. Future research should conduct an exploratory factor analysis to further explore potential other models.

Second, all CATI-R subscales as well as the social and non-social bifactors were closely related to each other. These correlations were not only stronger than in the original CATI validation study [8] but also more uniform, indicating that our revisions were consistent with capturing autistic traits as a multifaceted but single concept.

Third, the CATI-R demonstrated excellent internal consistency when assessed for the total scale and separate subscales. Indeed, internal consistency for the CATI-R appears to be stronger than for the original CATI, AQ,[5] and BAPQ for the current sample [87]. Likewise, the intercorrelation of the CATI-R items was higher than the one of the AQ or the BAPQ.

Fourth, we provided evidence for the nomological validity of the CATI-R at the individual level, by anchoring it in an existing framework, showing its correlation with the AQ and the BAPQ as measures of the same or closely related constructs.

Fifth, differences between male and female participants were limited to two subscales, suggesting that the CATI-R did not lose its original sensitivity to female autistic traits. Males scored significantly higher than females on the COM subscale and females scored significantly higher than males on the SEN subscale. While this difference may be the result of our large sample size [101], stronger communication difficulties in males align with research indicating a greater propensity for female autistic people to engage in social activities and communication (e.g., [13, 14, 24–27, 31]) as well as with findings for other measures of autistic traits. Looking at the AQ, for instance, Belcher and colleagues found, that female participants were more likely to endorse items related to social skills and communication [16]. One reason for such a different gender profile in autistic people might be a different socialisation of girls and boys in Western societies. Gender stereotypes may suggest the need for girls to be more sociable and empathic than boys [102].

[5] We used the AQ as a comparison although, in previous research, evidence does not support the proposed factor structure of the AQ (for an overview see [5]). In clinical research and practice, including epidemiological and cross-cultural research on autism [99], the AQ remains worldwide one of the most frequently used inventories for screening autism related behaviours [100]

To some extent, they may even be more 'trained' in social interactions. Alternatively, females might be less willing to endorse items related to communication difficulties because that would be less socially acceptable. For the same reason, males might endorse items related to sensory sensitivity less. Thus, to some extent, score differences on these items may also represent gender-based differences in response patterns rather than true differences in the construct of autistic traits.

Looking at the difference between female, male, and non-binary participants, non-binary participants scored higher than female and male participants; both on the total scale and on all subscales. This is in line with research suggesting that a higher proportion of autistic people identify as non-binary [88, 103]. However, our non-binary sample was also relatively small. As such, these trends warrant further investigation, in larger, prospective studies. Nevertheless, our results on gender reported here (as well as results in the further analyses examining the CATI-R's predictive ability for discriminating between autistic and non-autistic participants based on self-report data; see our Open Science Framework page: https://osf.io/vt4d2/) emphasises the need to explicitly include non-binary people in autism research (see [104], for a discussion on under-representation of specific demographic groups in research on neurodivergence).

Finally, inventories of autistic traits, such as the AQ, can be gender biased (see [16]), leading to higher or lower scores on certain items in specific gender groups. Such measurement bias or non-invariance indicates that the construct assumed to be reflected in the respective measure (e.g., autistic traits) does not have an equivalent structure or meaning across gender groups. Measurement non-invariance becomes particularly problematic when questionnaire scores are used for clinical decisions (see [6]). Therefore, we conducted a measurement invariance analysis to evaluate whether observed scores are comparable across female and male test takers (see the Open Science Framework page of this project: https://osf.io/vt4d2/ for an analysis including non-binary participants) by sequentially imposing between-group equality constraints on factor loadings, indicator intercepts, residual variances, and means [105]. This analysis indicated that the same underlying construct in the same way was being measured across the two gender groups. Specifically, we demonstrated the CATI-R scores could be meaningfully interpreted and compared between males and females, with evidence for (1) configural invariance; (2) metric invariance; (3) scalar invariance; and (4) residual/strict invariance. Therefore, we can assume that, across genders, participants neither interpreted the CATI-R differently nor differed in their likelihood of endorsing an item, indicating that CATI-R scores can be meaningfully interpreted and compared between gender groups.

Limitations and future research

The data reported here provides initial evidence for the validity of the CATI-R as a measure of individual differences in autistic traits in the general population. However, we acknowledge that a measure's true validity must be established via an iterative process that requires various empirical approaches [106, 107] and a nomological network built across multiple studies, possibly conducted by different researchers [108, 109]. We, thus, invite researchers to continuously develop and validate this questionnaire and others in a participatory manner.

For instance, there would be value in assessing the measure's discriminant validity in future work confirming that the CATI-R does not correlate with measures that are supposed to measure distinct constructs. This includes investigating the questionnaire's psychometric distinguishability from the original CATI to demonstrate its incremental value [110]. It would also be interesting to further investigate the CATI-R's nomological/criterion-based validity by examining intercorrelations between CATI-R scores and other theoretically related constructs, such as the Big Five personality traits, which account for 70% of variance in autistic-trait scores [111] or constructs that might be mistaken for autism (e.g., ADHD, see [112]). Finally, order effects should be investigated by assessing the CATI-R's psychometric properties when the order of items is randomised, and its test–retest reliability (i.e., reliability over time).

Despite our extensive examination of gender differences on the CATI-R, several important future avenues for exploration remain. Specifically, future work should explore if gender differences are similar for autistic and non-autistic people and whether these differences are moderated by age. While our wide age range in the autistic group (18–63 years) and the non-autistic group (18–77 years) indicates that our findings were unlikely to be confounded by age or historical changes in autism diagnosis, this aspect should be further investigated, considering research showing how some autistic individuals may no longer meet diagnostic criteria, depending on when their original diagnosis was received [71, 72, 113]. Furthermore, the relatively small number of non-binary participants in our sample restricts the generalisability of our findings to this population. Although practically challenging, future studies should target this population more extensively to examine the measure's psychometric properties for each gender and diagnostic group.

As the data came from five English-speaking countries, it might not capture the cultural diversity of other

populations. Given that socio-cultural expectations for appropriate behaviour can differ across cultures, future studies should include samples with more distinct socio-cultural backgrounds to also assess the cross-cultural validity of the CATI-R. This validation is particularly important given that the CATI and the CATI-R contain several items that reflect on a person's social experiences and behaviours.

Moreover, the current validation did not comprehensively assess participants for other potential mental health conditions that may impact CATI-R responses. Prospective research is needed to understand how different mental health conditions impact performance on the CATI-R, including conditions known to impact related constructs (e.g., social cognition and communication), such as anxiety, depression and psychosis [114–116].

Finally, our exploratory analyses suggest that the CATI-R holds the potential to be used as a screening tool. However, autistic people in our sample—as in the original CATI study [8]—were classified as autistic on self-report data alone, and of these participants, 88% reported having received a formal diagnosis. Whilst this rate is higher than that in the original CATI study, future work examining the validity, sensitivity and specificity of the CATI-R as a screening tool requires a data collection protocol that can confirm the diagnosis of autistic people using gold standard measures at the time of scale completion (e.g., ADOS-2; [95]).

Insights gained from the participatory project

In collaboration with all contributors, our project devised a set of guiding principles that shaped the revision. These principles can be applied to future endeavours for developing or updating individual difference measures that are related to neurodevelopmental or psychiatric divergences (Table 14; see the Open Science Framework page of this project: https://osf.io/vt4d2/ for more detailed guidelines).

Personality style measures for subclinical neurodivergent traits that are accessible and minimally stigmatising, reduce their negative impact on the test taker and lead to less biased results. Likewise, inventories that capture autistic experiences in a way that is not biased by the test taker's gender, improve the inventories' performance. In research, gender-invariant inventories can help recruit more diverse samples. In clinical practice, these measures may inform the development of training programs that aim to enhance recognition and understanding of the different autism phenotypes, allowing more individuals to access relevant services.

In contrast to meticulously planned participatory research projects (e.g., the recent development of the Self Assessment of Autistic Traits; [117]), this participatory project was organically initiated by autistic community members upon completing the original CATI in research they participated in. The benefit of this organic process was that autistic collaborators were highly motivated and entirely shaped the focus and priorities of the project. However, a more systematic approach to engaging in a participatory design study (e.g., [118]) would have had the additional benefits of prospectively determining roles and contributions, and modes of participation. Given that the current project transitioned from private, simultaneous conversations with participants in a previous

Table 14 Guidelines for personality style measures for sublcinical neurodivergent traits

Guideline	Example
Respect	
Use a neurodiversity approach	'impaired' versus 'different'
Use gender-inclusive language	'he' versus 'they'
Think non-stereotypically	'I do not feel a desire for social interactions' versus 'I prefer social interactions to occur in certain ways'
Accept language preferences of the community	'person with autism' versus 'autistic person'
Differentiate between in-group and out-group interaction partners	'Other people tend to misunderstand me' versus 'Non-autistic people tend to misunderstand me'
Consider positive aspects of neurodiversity	authenticity, honesty, loyalty, and deep focus
Involve the community in all its diversity	
Accessibility	
Be literal	'I rely on scripts when I talk with others' versus 'I plan how I will interact with others'
Be precise	'often' versus 'five times a day'
Be broad enough but not too general	Give at least two examples
Do not assume awareness of strategies	'I look for strategies to appear more sociable' versus 'I make an effort to appear more sociable'
Address past and present behaviour	Add instructions like 'Think about yourself both now and across your life'
Give additional room for explanation	Add optional free-text comments

study to a legitimate research endeavour, we grappled for instance, with how to acknowledge and recognise all of the contributions made by informants when the project rapidly but unexpectedly developed. In our study, collaboration and codesign involved neurodivergent and neurotypical researchers partnering with lay community members in our team. The modes of participation varied and included consultation (i.e. we sought their input), collaborative (i.e., we worked together on specific aspects of the project only, such as item revision), and collegiate (i.e., we worked together on all aspects of the research project, including data analysis). For future similar projects, we would consider the following aspects vital:

- clearly define the decision-making method before the start of the project,
- find roles and methods of engagement that fit everyone,
- offer opportunities for collaborators to engage in all stages of the research process (problem identification, research question development, data collection, interpretation of results),
- apply for funding to pay the non-academic collaborators, and
- embrace the heterogeneity of the autism community [69, 70] by including more diverse autistic people, including non-verbal individuals who have been widely excluded from research [119] instead of their parents or caregivers (for an exploration of how to address this challenge, see [120, 121]).

Conclusion

The current participatory project aimed to create an inventory of autistic traits that is not only empirically supported but also relevant to and accessible for autistic people. To this end, autistic and non-autistic people revised the CATI, an existing inventory of subclinical autistic traits, and applied a collective set of tests to provide preliminary evidence for the validity and reliability of the revised inventory. In sum, we argue that the CATI-R demonstrates substantial evidence for its validity as a self-report measure of autistic traits in the general population, whilst also capturing the experience of autistic people in accessible and non-stigmatising language. Moreover, our validation study indicates that the CATI-R demonstrates satisfactory measure invariance across genders. However, our findings—particularly involving non-binary participants—highlight the need for more research examining autistic traits across different gender profiles. Finally, this project demonstrated that research collaborations between neurotypical and neurodivergent people can foster shared learning, challenge stereotypes and advance autism research.

Abbreviations

AIC	Akaike information criterion
AQ	Autism-Spectrum Quotient
ANOVA	Analysis of variance
BAPQ	Broad Autism Phenotype Questionnaire
CAM	Camouflage (CATI subscale)
CATI	Comprehensive autistic trait inventory
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CI	Confidence interval
COM	Communication Difficulty (CATI & CATI-R subscale)
df	Degrees of freedom
EFA	Exploratory factor analysis
M	Mean
MAS	Masking (CATI-R subscale)
REP	Repetitive Behaviours (CATI & CATI-R subscale)
RIG	Cognitive Rigidity (CATI & CATI-R subscale)
RMSEA	Root mean square error of approximation
SD	Standard deviation
SEN	Sensory Sensitivity (CATI & CATI-R subscale)
SOC	Social Interactions (CATI & CATI-R subscale)
SRMR	Standardised root mean square residual
SRS	Social Responsiveness Scale
TLI	Tucker–Lewis index

Acknowledgements

Measure revisions were derived in collaboration with Hanan Bayrouti, Anna Dellarosa, Franca Dellarosa, Christian Stephan Erdmann Ambrosius Fischer, Wendy Higgins, Sharon Wieries, and many others who did not wish to be named.

Author contributions

FCH: Conceptualisation, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft, writing—review and editing, visualisation, supervision, project administration, funding acquisition. OT: Conceptualisation, methodology, validation, formal analysis, resources, data curation, writing—Original draft, writing—review and editing, visualisation, supervision, project administration, funding acquisition. NW: Methodology, validation, formal analysis, writing—review and editing. FF: Conceptualisation, investigation. BK: Conceptualisation, investigation. MaMa: Conceptualisation, investigation. MeMi: Conceptualisation, investigation. NC: Conceptualisation, methodology, validation, formal analysis, resources, data curation, writing—original draft, writing—review and editing, visualisation, supervision, project administration.

Funding

FCH received a PhD scholarship from Hans-Boeckler-Stiftung and the research was supported by IDEALAB and University of Potsdam.

availability of data and materials

The datasets generated and analysed during the current study, the analysis script, the revised inventory, and additional documentation of the revisions made are included in the Open Science Framework repository: https://osf.io/vt4d2/.

Declaration

Competing interests

The authors declare no competing interests.

Author details

[1]Macquarie University, Balaclava Rd, Macquarie Park, NSW 2109, Australia. [2]Potsdam University, Building 14, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany. [3]Flinders University, Sturt Rd, Bedford Park, SA 5042, Australia. [4]Berlin, Germany. [5]Krummhörn, Germany. [6]Bochum, Germany. [7]Sydney, Australia.

Received: 1 July 2024 Accepted: 12 January 2025
Published online: 06 February 2025

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. 5th ed. Arlington: American Psychiatric Association Arlington; 2013.
2. Kenny L, Hattersley C, Molins B, Buckley C, Povey C, Pellicano E. Which terms should be used to describe autism? Perspectives from the UK autism community. Autism Int J Res Pract. 2016;4(20):442–62. https://doi.org/10.1177/1362361315588200.
3. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health. 2017;5:307. https://doi.org/10.3389/fpubh.2017.00307.
4. Taylor LJ, Eggleston MJ, Thabrew H, van der Meer L, Waddington H, Whitehouse AJ, et al. An investigation of adherence to best practice guidelines for autism diagnosis in New Zealand. Autism. 2021;25(7):2087–100. https://doi.org/10.1177/13623613211015757.
5. Jia R, Steelman ZR, Jia HH. Psychometric assessments of three self-report autism scales (AQ, RBQ-2A, and SQ) for general adult populations. J Autism Dev Disord. 2019;49(5):1949–65. https://doi.org/10.1007/s10803-019-03880-x.
6. Gonzalez O, Pelham WEI. When does differential item functioning matter for screening? A method for empirical evaluation. Assessment. 2021;28(2):446–56. https://doi.org/10.1177/1073191120913618.
7. Lewis LF. Exploring the experience of self-diagnosis of autism spectrum disorder in adults. Arch Psychiatr Nurs. 2016;30(5):575–80. https://doi.org/10.1016/j.apnu.2016.03.009.
8. English M, Gignac GE, Visser TAW, Whitehouse A, Enns JT, Maybery M. The comprehensive autistic trait inventory (CATI): development and validation of a new measure of autistic traits in the general population. Mol Autism. 2021;12:37. https://doi.org/10.1186/s13229-021-00445-7.
9. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J Autism Dev Disord. 2001;31:5–7. https://doi.org/10.1023/A:1005653411471.
10. Hurley RSE, Losh M, Parlier M, Reznick JS, Piven J. The broad autism phenotype questionnaire. J Autism Dev Disord. 2007;37:1679–90. https://doi.org/10.1007/s10803-006-0299-3.
11. Bölte S. Brief report: The social responsiveness scale for adults (SRS-a): initial results in a German cohort. J Autism Dev Disord. 2011;42:1998–9. https://doi.org/10.1007/s10803-011-1424-5.
12. Strang JF, van der Miesen AI, Caplan R, Hughes C, daVanport S, Lai MC. Both sex- and gender-related factors should be considered in autism research and clinical practice. Autism. 2020;24(3):539–43. https://doi.org/10.1177/1362361320913192.
13. Hull L, Mandy W. Protective effect or missed diagnosis? Females with autism spectrum disorder. Future Neurol. 2017;12(3):159–69. https://doi.org/10.2217/fnl-2017-0006.
14. Lai MC, Lombardo M, Auyeung B, Chakrabarti B, Baron-Cohen S. Sex/gender differences and autism: setting the scene for future research. J Am Acad Child Adolesc Psychiatry. 2015;54(1):11–24. https://doi.org/10.1016/j.jaac.2014.10.003.
15. Lai MC, Szatmari P. Sex and gender impacts on the behavioural presentation and recognition of autism. Curr Opin Psychiatry. 2020;33(2):117–23. https://doi.org/10.1097/yco.0000000000000575.
16. Belcher HL, Uglik-Marucha N, Vitoratou S, Ford RM, Morein-Zamir S. Gender bias in autism screening: measurement invariance of different model frameworks of the autism spectrum quotient. BJPsych Open. 2023;9: e173. https://doi.org/10.1192/bjo.2023.562.
17. D'Mello AM, Frosch IR, Li CE, Cardinaux AL, Gabrieli JDE. Exclusion of females in autism research: empirical evidence for a 'leaky' recruitment-to-research pipeline. Autism Res. 2022;15(10):1929–40. https://doi.org/10.1002/aur.2795.
18. Rutter M, Caspi A, Moffitt TE. Using sex differences in psychopathology to study causal mechanisms: unifying issues and research strategies: using sex differences in psychopathology to study causal mechanisms. J Child Psychol Psychiatry. 2003;44(8):1092–115. https://doi.org/10.1111/1469-7610.00194.
19. Watkins EE, Zimmermann ZJ, Poling A. The gender of participants in published research involving people with autism spectrum disorders. Res Autism Spectr Disord. 2014;8(2):143–6. https://doi.org/10.1016/j.rasd.2013.10.010.
20. Dean M, Harwood R, Kasari C. The art of camouflage: gender differences in the social behaviors of girls and boys with autism spectrum disorder. Autism Int J Res Pract. 2017;21(6):678–89. https://doi.org/10.1177/1362361316671845.
21. Bargiela S, Steward R, Mandy W. The experiences of late-diagnosed women with autism spectrum conditions: an investigation of the female autism phenotype. J Autism Dev Disord. 2016;46(10):3281–94. https://doi.org/10.1007/s10803-016-2872-8.
22. Ridley R. Some difficulties behind the concept of the 'extreme male brain' in autism research. A theoretical review. Res Autism Spectr Disord. 2019;57:19–27. https://doi.org/10.1016/j.rasd.2018.09.007.
23. Wood-Downie H, Wong B, Kovshoff H, Cortese S, Hadwin JA. Research review: A systematic review and meta-analysis of sex/gender differences in social interaction and communication in autistic and nonautistic children and adolescents. J Child Psychol Psychiatry. 2021;62(8):922–36. https://doi.org/10.1111/jcpp.13337.
24. Hiller RM, Young RL, Weber N. Sex differences in autism spectrum disorder based on DSM-5 criteria: evidence from clinician and teacher reporting. J Abnorm Child Psychol. 2014;42(8):1381–93. https://doi.org/10.1007/s10802-014-9881-x20.
25. Hsiao MN, Tseng WL, Huang HY, Gau SSF. Effects of autistic traits on social and school adjustment in children and adolescents: the moderating roles of age and gender. Res Dev Disabil. 2013;34(1):254–65. https://doi.org/10.1016/j.ridd.2012.08.001.
26. Rynkiewicz A, Schuller B, Marchi E, Piana S, Camurri A, Lassalle A, et al. An investigation of the "female camouflage effect" in autism using a computerized ADOS-2 and a test of sex/gender differences. Mol Autism. 2016;7(10):10. https://doi.org/10.1186/s13229-016-0073-0.
27. Sedgewick F, Hill V, Yates R, Pickering L, Pellicano E. Gender differences in the social motivation and friendship experiences of autistic and nonautistic adolescents. J Autism Dev Disord. 2016;46(4):1297–306. https://doi.org/10.1007/s10803-015-2669-1.
28. Duvekot J, van der Ende J, Verhulst FC, Slappendel G, van Daalen E, Maras A, et al. Factors influencing the probability of a diagnosis of autism spectrum disorder in girls versus boys. Autism Int J Res Pract. 2017;21(6):646–58. https://doi.org/10.1177/1362361316672178.
29. Halladay AK, Bishop S, Constantino JN, Daniels AM, Koenig K, Palmer K, et al. Sex and gender differences in autism spectrum disorder: summarizing evidence gaps and identifying emerging areas of priority. Mol Autism. 2015;6(1):36–6. https://doi.org/10.1186/s13229-015-0019-y.
30. Lai MC, Lombardo M, Ruigrok A, Chakrabarti B, Auyeung B, Szatmari P, et al. Quantifying and exploring camouflaging in men and women with autism. Autism. 2017;21(6):690–702. https://doi.org/10.1177/1362361316671012.
31. Cho S, Cola M, Knox A, Pelella MR, Russell A, Hauptmann A, et al. Sex differences in the temporal dynamics of autistic children's natural conversations. Mol Autism. 2023;14(1):13. https://doi.org/10.1186/s13229-023-00545-6.
32. Mandy W, Chilvers R, Chowdhury U, Salter G, Seigal A, Skuse D. Sex differences in autism spectrum disorder: evidence from a large sample of children and adolescents. J Autism Dev Disord. 2012;42(7):1304–13. https://doi.org/10.1007/s10803-011-1356-0.
33. Huke V, Turk J, Saeidi S, Kent A, Morgan JF. Autism spectrum disorders in eating disorder populations: a systematic review. Autism Spectr disord Eat Disor Eur Eat Disor Rev. 2013;21(5):345–51. https://doi.org/10.1002/erv.2244.
34. Hull L, Lai MC, Baron-Cohen S, Allison C, Smith P, Petrides K, et al. Gender differences in self-reported camouflaging in autistic and non-autistic adults. Autism Int J Res Pract. 2020;24(2):352–63. https://doi.org/10.1177/1362361319864804.
35. Hull L, Petrides KV, Allison C, Smith P, Baron-Cohen S, Lai MC, et al. "Putting on my best normal": social camouflaging in adults with autism spectrum conditions. J Autism Dev Disord. 2017;47(8):2519–34. https://doi.org/10.1007/s10803-017-3166-5.

36. Pearson A, Rose K. A conceptual analysis of autistic masking: understanding the narrative of stigma and the illusion of choice. Autism Adulthood. 2021;3(1):52–60. https://doi.org/10.1089/aut.2020.0043.
37. Dworzynski K, Ronald A, Bolton P, Happé F. How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? J Am Acad Child Adolesc Psychiatry. 2012;51(8):788–97. https://doi.org/10.1016/j.jaac.2012.05.018.
38. Belcher HL, Morein-Zamir S, Stagg SD, Ford RM. Shining a light on a hidden population: social functioning and mental health in women reporting autistic traits but lacking diagnosis. J Autism Dev Disord. 2023;53(8):3118–32. https://doi.org/10.1007/s10803-022-05583-2.
39. Hull L, Petrides K, Mandy W. The female autism phenotype and camouflaging: a narrative review. Review J Autism Dev Disord. 2020;7:306–17. https://doi.org/10.1007/s40489-020-00197-9.
40. Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, et al. Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2014. MMWR Surveill Summ. 2018;67(6):1–23. https://doi.org/10.15585/mmwr.ss6706a1.
41. Wilson C, Murphy CM, McAlonan G, Robertson DM, Spain D, Hayward H, et al. Does sex influence the diagnostic evaluation of autism spectrum disorder in adults? Autism Int J Res Pract. 2016;20(7):808–19. https://doi.org/10.1177/1362361315611381.
42. Kreiser Wells N, White S. ASD in females: Are we overstating the gender difference in diagnosis? Clin Child Fam Psychol Rev. 2013;17:67–84. https://doi.org/10.1007/s10567-013-0148-9.
43. Lockwood Estrin G, Milner V, Spain D, Happe F, Colvert E. Barriers to autism spectrum disorder diagnosis for young women and girls: a systematic review. Rev J Autism Dev Disord. 2021;8:1–17. https://doi.org/10.1007/s40489-020-00225-8.
44. Begeer S, Mandell D, Wijnker-Holmes B, Venderbosch S, Rem D, Stekelenburg F, et al. Sex differences in the timing of identification among children and adults with autism spectrum disorders. J Autism Dev Disord. 2013;43(5):1151–6. https://doi.org/10.1007/s10803-012-1656-z.
45. Gesi C, Migliarese G, Torriero S, Capellazzi M, Omboni AC, Cerveri G, et al. Gender differences in misdiagnosis and delayed diagnosis among adults with autism spectrum disorder with no language or intellectual disability. Brain Sci. 2021;11(7):912. https://doi.org/10.3390/brainsci11070912.
46. Whitlock A, Fulton K, Lai M, Pellicano E, Mandy W. Recognition of girls on the autism spectrum by primary school educators: an experimental study. Autism Res. 2020;13(8):1358–72. https://doi.org/10.1002/aur.2316.
47. Pellicano E, Dinsmore A, Charman T. What should autism research focus upon? Community views and priorities from the United Kingdom. Autism Int J Res Pract. 2014;18(7):756–70. https://doi.org/10.1177/1362361314529627.
48. Belek B. Articulating sensory sensitivity: from bodies with autism to autistic bodies. Med Anthropol. 2019;38(1):30–43. https://doi.org/10.1080/01459740.2018.1460750.
49. Dinishak J. Autistic autobiography and hermeneutical injustice. Metaphilosophy. 2021;52(5):556–69. https://doi.org/10.1111/meta.12514.
50. Davis R, Crompton CJ. What do new findings about social interaction in autistic adults mean for neurodevelopmental research? Perspect Psychol Sci. 2021;16(3):649–53. https://doi.org/10.1177/1745691620958010.
51. Milton DEM. On the ontological status of autism: the 'double empathy problem'. Disabil Soc. 2012;27(6):883–7. https://doi.org/10.1080/09687599.2012.710008.
52. van Rentergem JAA, Lever AG, Geurts HM. Negatively phrased items of the autism spectrum quotient function differently for groups with and without autism. Autism. 2019;23(7):1752–64. https://doi.org/10.1177/1362361319828361.
53. Bottema-Beutel K, Kapp SK, Lester JN, Sasson NJ, Hand BN. Avoiding ableist language: suggestions for autism researchers. Autism Adulthood. 2021;3(1):18–29. https://doi.org/10.1089/aut.2020.0014.
54. Dwyer P. The neurodiversity approach (es): What are they and what do they mean for researchers? Hum Dev. 2022;66(2):73–92. https://doi.org/10.1159/000523723.
55. Jones SC. Measuring the wrong thing the right way? Time to rethink autism research tools. Autism Adulthood. 2022;4(2):104–9. https://doi.org/10.1089/aut.2021.0050.
56. Balazs CL, Morello-Frosch R. The three r's: how community based participatory research strengthens the rigor, relevance and reach of science. Environ Justice. 2013;6(1):9–16. https://doi.org/10.1089/env.2012.0017.
57. Gillespie-Lynch K, Bublitz D, Donachie A, Wong V, Brooks PJ, D'Onofrio J. "For a long time our voices have been hushed": using student perspectives to develop supports for neurodiverse college students. Front Psychol. 2017;8:544. https://doi.org/10.3389/fpsyg.2017.00544.
58. Gillespie-Lynch K, Bisson JB, Saade S, Obeid R, Kofner B, Harrison AJ, et al. If you want to develop an effective autism training, ask autistic students to help you. Autism Int J Res Pract. 2022;26(5):1082–94. https://doi.org/10.1177/13623613211041006.
59. den Houting J, Higgins J, Isaacs K, Mahony J, Pellicano E. 'I'm not just a guinea pig': academic and community perceptions of participatory autism research. Autism Int J Res Pract. 2021;25(1):148–63. https://doi.org/10.1177/1362361320951696.
60. Vaughn L, Jacquez F. Participatory research methods—choice points in the research process. J Particip Res Methods. 2020. https://doi.org/10.35844/001c.13244.
61. Pellicano E, den Houting J. Annual research review: shifting from 'normal science' to neurodiversity in autism science. J Child Psychol Psychiatry. 2022;63(4):381–96. https://doi.org/10.1111/jcpp.13534.
62. Pellicano E, Lawson W, Hall G, Mahony J, Lilley R, Heyworth M, et al. "I knew she'd get it, and get me": participants' perspectives of a participatory autism research project. Autism Adulthood. 2022;4(2):120–9. https://doi.org/10.1089/aut.2021.0039.
63. Poulsen R, Brownlow C, Lawson W, Pellicano E. Meaningful research for autistic people? Ask autistics! Autism. 2022;26(1):3–5. https://doi.org/10.1177/13623613211064421.
64. Morrison KE, DeBrabander KM, Jones DR, Ackerman RA, Sasson NJ. Social cognition, social skill, and social motivation minimally predict social interaction outcomes for autistic and non-autistic adults. Front Psychol. 2020;11: 591100. https://doi.org/10.3389/fpsyg.2020.591100.
65. Gillespie-Lynch K, Kapp S, Brooks P, Pickens J, Schwartzman B. Whose expertise is it? Evidence for autistic adults as critical autism experts. Front Psychol. 2017;8:438. https://doi.org/10.3389/fpsyg.2017.00438.
66. Crompton CJ, Ropar D, Evans-Williams CV, Flynn EG, Fletcher-Watson S. Autistic peer-to-peer information transfer is highly effective. Autism Int J Res Pract. 2020;24(7):1704–12. https://doi.org/10.1177/1362361320919286.
67. Edey R, Cook J, Brewer R, Johnson MH, Bird G, Press C. Interaction takes two: typical adults exhibit mind-blindness towards those with autism spectrum disorder. J Abnorm Psychol (1965). 2016;125(7):879–85. https://doi.org/10.1037/abn0000199.
68. Rundblad G, Annaz D. The atypical development of metaphor and metonymy comprehension in children with autism. Autism. 2010;14(1):29–46. https://doi.org/10.1177/1362361309340667.
69. Bloch C, Viswanathan S, Tepest R, Jording M, Falter-Wagner CM, Vogeley K. Differentiated, rather than shared, strategies for time-coordinated action in social and non-social domains in autistic individuals. Cortex. 2023;166:207–32. https://doi.org/10.1016/j.cortex.2023.05.008.
70. Lai MC, Lombardo MV, Chakrabarti B, Baron-Cohen S. Subgrouping the autism "spectrum": reflections on DSM-5. PLoS Biol. 2013;11(4):1–7. https://doi.org/10.1371/journal.pbio.1001544.
71. Fein D, Barton M, Eigsti IM, Kelley E, Naigles L, Schultz RT, et al. Optimal outcome in individuals with a history of autism. J Child Psychol Psychiatry. 2013;54(2):195–205. https://doi.org/10.1111/jcpp.12037.
72. Waizbard-Bartov E, Ferrer E, Heath B, Rogers SJ, Nordahl CW, Solomon M, et al. Identifying autism symptom severity trajectories across childhood. Autism Res. 2022;15(4):687–701. https://doi.org/10.1002/aur.2674.
73. Barrett S, Uljarevic M, Jones C, Leekam S. Assessing subtypes of restricted and repetitive behaviour using the adult repetitive behaviour questionnaire-2 in autistic adults. Mol Autism. 2018;9:58. https://doi.org/10.1186/s13229-018-0242-4.
74. Livingston LA, Happé F. Conceptualising compensation in neurodevelopmental disorders: reflections from autism spectrum disorder.

Neurosci Biobehav Rev. 2017;80:729–42. https://doi.org/10.1016/j.neubiorev.2017.06.005.
75. Douglas BD, Ewell PJ, Brauer M. Data quality in online human-subjects research: comparisons between MTurk, prolific, CloudResearch, qualtrics, and SONA. PloS ONE. 2023;18(3): e0279720. https://doi.org/10.1371/journal.pone.0279720.
76. Peer E, Rothschild D, Gordon A, Evernden Z, Damer E. Data quality of platforms and panels for online behavioral research. Behav Res Methods. 2022;54(4):1643–62. https://doi.org/10.3758/s13428-021-01694-3.
77. Stevenson JL, Hart KR. Psychometric properties of the autism-spectrum quotient for assessing low and high levels of autistic traits in college students. J Autism Dev Disord. 2017;47(6):1838–53. https://doi.org/10.1007/s10803-017-3109-1.
78. Kline RB. Principles and practice of structural equation modeling. 4th ed. Methodology in the social sciences. New York: Guilford Publications; 2015.
79. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Softw. 2012;48(2):1–36. https://doi.org/10.18637/jss.v048.i02.
80. Gignac GE. The higher-order model imposes a proportionality constraint: that is why the bifactor model tends to fit better. Intelligence. 2016;55:57–68. https://doi.org/10.1016/j.intell.2016.01.006.
81. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res. 2004;33(2):261–304. https://doi.org/10.1177/0049124104268644.
82. Tabachnick BG, Fidell LS. Experimental designs using ANOVA. Duxbury applied series. Pacific Grove: Thomson/Brooks/Cole; 2007.
83. Comrey AL, Lee HB. A first course in factor analysis. Milton Park: Taylor & Francis; 2013.
84. Briggs SR, Cheek JM. The role of factor analysis in the development and evaluation of personality scales. J Personal. 1986;54(1):106–48. https://doi.org/10.1111/j.1467-6494.1986.tb00391.x.
85. Gignac GE. On the inappropriateness of using items to calculate total scale score reliability via coefficient alpha for multidimensional scales. Eur J Psychol Assessm Off Organ Eur Assoc Psychol Assess. 2014;30(2):130–9. https://doi.org/10.1027/1015-5759/a000181.
86. Cronbach LJ, Schönemann P, McKie D. Alpha coefficients for stratified-parallel tests. Educ Psychol Meas. 1965;25(2):291–312. https://doi.org/10.1177/001316446502500201.
87. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ (Aust Sci Educ Res Assoc). 2018;48(6):1273–96. https://doi.org/10.1007/s11165-016-9602-2.
88. Mahfouda S, Panos C, Whitehouse A, Thomas C, Maybery M, Strauss P, et al. Mental health correlates of autism spectrum disorder in gender diverse young people: evidence from a specialised child and adolescent gender clinic in australia. J Clin Med. 2019. https://doi.org/10.3390/jcm8101503.
89. Warrier V, Greenberg DM, Weir E, Buckingham C, Smith P, Lai MC, et al. Elevated rates of autism, other neurodevelopmental and psychiatric diagnoses, and autistic traits in transgender and gender-diverse individuals. Nat Commun. 2020;11(1):3959–9. https://doi.org/10.1038/s41467-020-17794-1.
90. Kassambara A. rstatix: pipe-friendly framework for basic statistical tests [Internet]. 2023. Available from: https://CRAN.R-project.org/package=rstatix.
91. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. J Stat Softw. 2017;82(13):1–26.
92. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev Rev. 2016;41:71–90. https://doi.org/10.1016/j.dr.2016.06.004.
93. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Model Multidiscip J. 2007;14(3):464–504. https://doi.org/10.1080/10705510701301834.
94. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model. 2002;9(2):233–55. https://doi.org/10.1207/S15328007SEM0902D_5.
95. Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Pickles A. Autism diagnostic observation schedule (ADOS-2). 2nd ed. Los Angeles: Western Psychological Services; 2012.
96. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model. 1999;6(1):1–55. https://doi.org/10.1080/10705519909540118.
97. Guadagnoli E, Velicer W. Relation of sample size to the stability of component patterns. Psychol Bull. 1988;103:265–75. https://doi.org/10.1037/0033-2909.103.2.265.
98. Benevides TW, Shore SM, Palmer K, Duncan P, Plank A, Andresen ML, et al. Listening to the autistic voice: mental health priorities to guide research and practice in autism from a stakeholder-driven project. Autism. 2020;24(4):822–33. https://doi.org/10.1177/1362361320908410.
99. Wakabayashi A, Uchiyama T, Tojo Y, Yoshida Y, Kuroda M, Baron-Cohen S, et al. The autsim-spectrum quotient (AQ) japanese children's version: comparison between high-functioning children with autism spectrum disorders and normal controls. Shinrigaku kenkyū. 2007;77(6):534–40. https://doi.org/10.1007/s10803-006-0181-3.
100. Broderick N, Wade JL, Meyer JP, Hull M, Reeve RE. Model invariance across genders of the broad autism phenotype questionnaire. J Autism Dev Disord. 2015;45(10):3133–47. https://doi.org/10.1007/s10803-015-2472-z.
101. Bakker A, Cai J, English L, Kaiser G, Mesa V, Van Dooren W. Beyond small, medium, or large: points of consideration when interpreting effect sizes. Educ Stud Math. 2019;102(1):1–8. https://doi.org/10.1007/s10649-019-09908-4.
102. Krahn TM, Fenton A. The extreme male brain theory of autism and the potential adverse effects for boys and girls with autism. J Bioethical Inq. 2012;9(1):93–103. https://doi.org/10.1007/s11673-011-9350-y.
103. Kallitsounaki A, Williams DM. Autism spectrum disorder and gender dysphoria/incongruence. A systematic literature review and meta-analysis. J Autism Dev Disord. 2023;53(8):3103–17. https://doi.org/10.1007/s10803-022-05517-y.
104. Bölte S, Neufeld J, Marschik PB, Williams ZJ, Gallagher L, Lai MC. Sex and gender in neurodevelopmental conditions. Nat Rev Neurol. 2023;19(3):136–59. https://doi.org/10.1038/s41582-023-00774-6.
105. Meredith W. Measurement invariance, factor analysis and factorial invariance. Psychometrika. 1993;58(4):525–43. https://doi.org/10.1007/BF02294825.
106. Chang H. Inventing temperature: measurement and scientific progress. Oxford studies in the philosophy of science. New York: Oxford University Press; 2004.
107. Clark LA, Watson D. Constructing validity: new developments in creating objective measuring instruments. Psychol Assess. 2019;31(12):1412–27. https://doi.org/10.1037/pas0000626.
108. Trafimow D. Construct validity. Int J Aviat Res. 2022;14(01):96–110.
109. Truijens FL, Desmet M, De Coster E, Uyttenhove H, Deeren B, Meganck R. When quantitative measures become a qualitative storybook: a phenomenological case analysis of validity and performativity of questionnaire administration in psychotherapy research. Qual Res Psychol. 2022;19(1):244–87. https://doi.org/10.1080/14780887.2019.1579287.
110. Murphy BA, Hall JA, Duong F. It looks like construct validity, but look again: comment on Clutterbuck et al. (2021) and recommendations for test developers in the broad "empathy" domain. Psychol Assess. 2022;34(4):397–404. https://doi.org/10.1037/pas0001063.
111. Schwartzman BC, Wood JJ, Kapp SK. Can the five factor model of personality account for the variability of autism symptom expression? Multivariate approaches to behavioral phenotyping in adult autism spectrum disorder. J Autism Dev Disord. 2016;46(1):253–72. https://doi.org/10.1007/s10803-015-2571-x.
112. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. J Am Acad Child Adolesc Psychiatry. 2008;47(8):921–9. https://doi.org/10.1097/CHI.0b013e318179964f.
113. Helles A, Gillberg IC, Gillberg C, Billstedt E. Asperger syndrome in males over two decades: quality of life in relation to diagnostic stability and psychiatric comorbidity. Autism. 2017;21(4):458–69. https://doi.org/10.1177/1362361316650090.
114. Dall M, Fellinger J, Holzinger D. The link between social communication and mental health from childhood to young adulthood: a systematic review. Front Psychiatry. 2022;13: 944815.
115. Shilyansky C, Williams LM, Gyurak A, Harris A, Usherwood T, Etkin A. Effect of antidepressant treatment on cognitive impairments associated

with depression: A randomised longitudinal study. Lancet Psychiatry. 2016;3(5):425–35. https://doi.org/10.1016/S2215-0366(16)00012-2.
116. Çıkrıkçı N. Explaining association between interpersonal communication competence and depression through need satisfaction, anxiety, and stress. Curr Psychol. 2024;43(31):25468–80. https://doi.org/10.1007/s12144-024-06250-8.
117. Ratto AB, Bascom J, daVanport S, Strang JF, Anthony LG, Verbalis A, et al. Centering the inner experience of autism: development of the self-assessment of autistic traits. Autism Adulthood. 2023;5(1):93–105. https://doi.org/10.1089/aut.2021.0099.
118. Duea SR, Zimmerman EB, Vaughn LM, Dias S, Harris J. A guide to selecting participatory research methods based on project and partnership goals. J Particip Res Methods. 2022;3(1):1–28. https://doi.org/10.35844/001c.32605.
119. Russell G, Kapp SK, Elliott D, Elphick C, Gwernan-Jones R, Owens C. Mapping the autistic advantage from the accounts of adults diagnosed with autism: a qualitative study. Autism Adulthood. 2019;1(2):124–33. https://doi.org/10.1089/aut.2018.0035.
120. Hills KJ, Clapton J, Dorsett P, Andersen K. Conducting research with people with nonverbal autism: an inclusive methodological approach. J Soc Incl. 2020;11(1):5. https://doi.org/10.36251/josi.162.
121. Ridout S. The autistic voice and creative methodologies. Qual Res J. 2017;17:52–64. https://doi.org/10.1108/QRJ-07-2016-0046.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.