# Two validation studies of a performance validity test for autistic adults

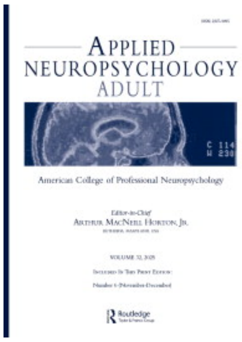
Scheeren, Anke M.; Olde Dubbelink, Linda; Lever, Anne Geeke; Geurts, Hilde M.

**DOI**
10.1080/23279095.2024.2305206

**Publication date**
2025

**Document Version**
Final published version

**Published in**
Applied Neuropsychology: Adult



Applied Neuropsychology: Adult

ISSN: 2327-9095 (Print) 2327-9109 (Online) Journal homepage: www.tandfonline.com/journals/hapn21

# Two validation studies of a performance validity test for autistic adults

Anke M. Scheeren, Linda Olde Dubbelink, Anne Geeke Lever & Hilde M. Geurts

**To cite this article:** Anke M. Scheeren, Linda Olde Dubbelink, Anne Geeke Lever & Hilde M. Geurts (2025) Two validation studies of a performance validity test for autistic adults, Applied Neuropsychology: Adult, 32:6, 1698-1710, DOI: 10.1080/23279095.2024.2305206

**To link to this article:** https://doi.org/10.1080/23279095.2024.2305206

APPLIED NEUROPSYCHOLOGY: ADULT
2025, VOL. 32, NO. 6, 1698–1710
https://doi.org/10.1080/23279095.2024.2305206

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

# Two validation studies of a performance validity test for autistic adults

Anke M. Scheeren[a], Linda Olde Dubbelink[a*], Anne Geeke Lever[a#] and Hilde M. Geurts[a,b]

[a]Dutch Autism & ADHD Research Center, Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands;
[b]Dr. Leo Kannerhuis, autism clinic, Amsterdam, the Netherlands

**ABSTRACT**

In two studies we examined the potential of a simple emotion recognition task, the Morel Emotional Numbing Test (MENT), as a performance validity test (PVT) for autism-related cognitive difficulties in adulthood. The aim of a PVT is to indicate non-credible performance, which can aid the interpretation of psychological assessments. There are currently no validated PVTs for autism-related difficulties in adulthood. In Study 1, non-autistic university students (aged 18–46 years) were instructed to simulate that they were autistic during a psychological assessment (simulation condition; $n = 26$). These students made more errors on the MENT than those instructed to do their best (control condition; $n = 26$). In Study 2, we tested how well autistic adults performed on the MENT. We found that clinically diagnosed autistic adults and non-autistic adults (both $n = 25$; 27–57 years; IQ > 80) performed equally well on the MENT. Moreover, autistic adults made significantly fewer errors than the instructed simulators in Study 1. The MENT reached a specificity of ≥98% (identifying 100% of non-simulators as non-simulator in Study 1 and 98% in Study 2) and a sensitivity of 96% (identifying 96% of simulators as simulator). Together these findings provide the first empirical evidence for the validity of the MENT as a potential PVT for autism-related cognitive difficulties.

**KEYWORDS**
Adulthood; autism; exaggeration; MENT; performance validity test

Autism was typically known as a diagnosis which people received in childhood, yet, there is a growing consensus among clinicians and researchers that autism can also be diagnosed in adulthood (e.g., Lai & Baron-Cohen, 2015). However, clinicians and researchers screening for autistic characteristics in adults face several unique challenges, including a relatively small empirical knowledge base of autism in adulthood, a lack of reliable information on a person's developmental history (due to deceased or old parents), and a restricted number of age-appropriate diagnostic instruments that can reliably discriminate between autism and mental health conditions (Bal & Taylor, 2019; Wigham et al., 2019). These factors all increase the risk of wrong diagnoses (Fombonne, 2023), including false negatives (missed autism diagnosis) and false positives (false autism diagnosis). Valid and reliable assessment tools are therefore crucial to increase the validity of an autism diagnosis in adulthood.

In the field of neuropsychology, symptom validity tests (SVTs; assessing the validity of symptoms) and performance validity tests (PVTs; assessing the validity of test performance) are widely used to ascertain the "accuracy or truthfulness of the examinee's behavioral presentation (signs), self-reported symptoms (including their cause and course), or performance on neuropsychological measures" (Bush et al., 2005, p. 420). People may exaggerate or even simulate particular symptoms or difficulties in order to receive an incentive such as financial and medical aids (i.e., specific drugs), access to treatment or support, exemptions, or reduced responsibility for criminal actions (Hardeman, 2006; Maras et al., 2019; Merckelbach et al., 2009). Although the exact prevalence of exaggeration and simulation of psychological or physical difficulties is difficult to determine, and also varies according to context (e.g., criminal/health care/vocational/legal; Sweet et al., 2021), sample's characteristics, and SVTs/PVTs used (Dandachi-FitzGerald et al., 2013), exaggeration/simulation is estimated to be as high as 30–60% for cases of whiplash, chronic fatigue, and chronic toxic encephalopathy (e.g., Hardeman, 2006; Mestdagh et al., 2019). For mental health conditions such as Attention Deficit Hyperactivity Disorder (ADHD; Sullivan et al., 2007), depression (Monaro et al., 2018), and Post Traumatic Stress Disorders (PTSD; Young, 2017) the prevalence estimations range from 10% to almost 50%. As yet, SVTs and PVTs are not commonly used in the diagnostic screening for autism

**CONTACT** Anke M. Scheeren A.M.Scheeren@vu.nl VU University Amsterdam, Faculty of Behavioural and Movement Sciences, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/23279095.2024.2305206.
*Present address: GGNET, mental health care center, Doetinchem, the Netherlands.
#Present address: Dimence, mental health care center, Deventer, the Netherlands.

in adults, even though a recent survey among clinical autism professionals showed that about half suspected feigning of autism in one or more adults (Murphy & Broyd, 2023). Findings of one earlier study (Dandachi-FitzGerald et al., 2011) suggest that up to 16% of autistic adults might over-report symptoms of psychopathology and underperform on a memory-related PVT. However, both measures used are not key to an autism diagnosis, making it difficult to interpret as an SVT/PVT for autism.

Good quality SVTs and PVTs would increase the validity of an autism diagnosis in adulthood by reducing false positives (people falsely diagnosed with autism). Failing to distinguish between genuine and exaggerated/simulated difficulties has implications on an individual and population level. First, a false autism diagnosis may give a person access to types of care or support that do not fit the person's needs and abilities, thereby prolonging the route to adequate care in case of mental health problems unrelated to autism. Moreover, a false diagnosis likely puts other people (true positives and false negatives) at a disadvantage due to limited aids (finances, care). Invalid diagnoses also lead to unwanted "noise" in research data, decreasing potential group differences between autistic and non-autistic participants as well as the effectiveness of autism-specific interventions (Fombonne, 2023). Recently, this problem has been highlighted in online qualitative autism research, where some participants were suspected of posing as autistic, presumably for financial reasons (Pellicano et al., 2023). Furthermore, invalid diagnoses make it difficult to detect a specific pattern of cognitive strengths and weaknesses related to a specific diagnosis (Hoelzle et al., 2019). Finally, invalid diagnoses potentially endanger society's trust in the existence of a specific diagnosis. This is why it is relevant to develop and assess an SVT/PVT for autism in adults.

Even though autism in adults may still be underdiagnosed (Nylander et al., 2013), there are several reasons to suppose that, at least in some cases, exaggerating or simulating autism symptoms or autism-related difficulties may be (a) possible and (b) beneficial. First, we argue it is *possible* because in autism diagnostic assessments, like in many other mental health conditions, the main informant is the adult themself. While it is recommended (NICE, 2012) to gather information from a proxy and to use standardized diagnostic measures, clinicians often rely on self-report and their own impressions (e.g., Taylor et al., 2016). Information via proxies may be inaccessible (e.g., due to old age, death of parents) and known standardized diagnostic measures (e.g., Autism Diagnostic Observation Scale; Lord et al., 2000, 2012) are not always used, as they are considered to be suboptimal (Shulman et al., 2020; Timimi et al., 2019), less suitable for adults, and time/money consuming. Second, a false autism diagnosis might be *beneficial* for some adults, as it could increase their and others' understanding and acceptance of certain behaviors, choices or preferences (e.g., sticking to routines) compared to when no diagnosis or when an other psychiatric diagnosis is provided. Indeed, even though people with autism may experience stigmatization and ableism (Cage et al., 2019), the stigma of autism appears substantially less compared to personality disorders, psychopathy, schizophrenia, or addiction (Feldman & Crandall, 2007; Gillespie-Lynch et al., 2021). This positive appraisal of autism relative to other conditions may be due to an increased emphasis on neurodiversity, strengths and positive aspects of being autistic (Gillespie-Lynch et al., 2017), which has successfully contributed to the crucial emancipation of autistic adults, but may (inadvertently) also have increased the risk of false positives. Additionally, in various health care policies an autism diagnosis gives access to or is a requirement to obtain social benefits, support, treatment, or care. In research, autistic participants may be financially compensated for their participation (Pellicano et al., 2023). In forensic cases, an autism diagnosis may lead to reduced accountability for criminal actions. As was shown in a vignette study, a disclosure of an autism diagnosis of the defendant resulted in fewer guilty verdicts and more lenient sentencing as compared to when no diagnostic label was provided (Maras et al., 2019). Given the possibility and the potential benefits of false (claims of) autism diagnoses, it is important for both clinical practice and research to reduce the number of false positives. Currently there are no tools designed specifically to target the validity of autistic symptoms (via SVTs) or autism-related (cognitive) challenges (via PVTs) in adults. Therefore, in two complementary studies we will assess the suitability of a simple emotion recognition task, the Morel Emotional Numbing Test (MENT; Morel, 1998), as a potential PVT for autism-related cognitive difficulties.

The MENT was originally developed as a PVT for PTSD (Morel, 1998). The test consists of pictures of prototypical facial expressions and emotion words. The participant is instructed to link an emotion word (e.g., afraid) to one of two easily recognizable facial expressions, and, vice versa, pair a facial expression with one of two emotion words. According to Hartman (2002), an effective PVT should meet three criteria: (1) it should lure a simulating individual to perform beneath their true level of ability, (2) it should be a face-valid measure of the condition of interest, and (3) the PVT should in fact be insensitive to the condition of interest. The MENT meets all of these criteria as a PVT for PTSD (Morel & Shepherd, 2008). The MENT may also be a suitable PVT for autism-related cognitive difficulties, because most studies show intact recognition of simple emotions in individuals with an autism diagnosis when based on photographs of prototypical whole-face expressions (Nuske et al., 2013), while (lay) people may falsely believe that autistic adults have serious difficulties in emotion recognition on all levels of complexity. Therefore, we expect individuals who exaggerate or simulate autism-related difficulties to make substantially more mistakes on the MENT compared to autistic adults (with a validated diagnosis), who may perform quite well on the MENT.

With two separate studies we examined whether the MENT meets the three criteria of a valid PVT for autism-related cognitive difficulties. The primary goal of Study 1 was to examine whether the MENT lures simulators into underperformance (criterion 1) and the secondary goal was to check whether the MENT appears to be a face-valid measure of autism (criterion 2). We compared MENT performances of non-autistic students who either received the instruction to simulate autism (simulation condition) or who were asked to perform to the best of their knowledge and ability (control condition). A simulation design is frequently used in an early, exploratory stage of research to examine the discriminatory power of new PVTs (Sweet et al., 2021). We hypothesized that the group of instructed simulators would perform more poorly on the MENT compared to controls. The primary goal of Study 2 was to examine insensitivity of the MENT to an autism diagnosis (criterion 3) and a secondary goal was to check the face validity of the MENT (criterion 2). We did so by comparing MENT performances of adults with a clinical autism diagnosis to a matched non-autistic control group. We hypothesized that the MENT is insensitive to an autism diagnosis, so autistic adults were expected to perform similarly as non-autistic controls on the MENT.

## Study 1 method

### Participants

Participants were 52 first year psychology and psychobiology students of the University of Amsterdam who received credits for participating in this study (in 2014–2015). They were randomly assigned to one of two conditions (26 in the simulation condition, 26 in the control condition). Participants in the two conditions were comparable with regard to age ($t(44) = 0.93$, $p = .36$, Cohen's $d = .27$), distribution of men and women ($\chi^2(1) = 0.92$, $p = .34$, $\varphi = .13$) and symptoms of depression ($t(50) = 0.86$, $p = .39$, Cohen's $d = .24$) based on the Beck Depression Inventory-II (Beck et al., 1996). Despite randomization, participants in the control condition tended to report slightly more autism characteristics on the Autism-Spectrum Quotient (AQ; Baron-Cohen et al., 2001), an autism trait questionnaire, than those in the simulation condition at screening ($t(50) = 1.91$, $p = .06$, Cohen's $d = .53$), please see Table 1 for the descriptives. However, none of the participants in either condition had or approached an AQ score of 32 or higher, indicating that no one showed clinically significant levels of autistic traits (recommended cutoff for a general population sample by Woodbury-Smith et al., 2005).

### Measures

The MENT was embedded in a complete (neuro)psychological test battery so as to create a more realistic context and to put less emphasis on the MENT. Even though a complete neuropsychological assessment is not a standard procedure when assessing autism in adults, it is common to use multiple tests, questionnaires and interviews.

#### Morel Emotional Numbing Test (MENT)

The MENT (Morel, 1998) was originally developed as a performance validity test for PTSD. It is a paper-and-pencil task consisting of three sets of emotion stimuli. The stimuli are pictures of intense whole-face emotional facial expressions of two young men (one Caucasian, one African-American) and words of emotions (happy, sad, angry, calm, surprised, afraid, confused, sleepy, frustrated, and shy). The emotion words in the current study were derived from the Dutch version of the MENT developed by Geraerts et al. (2007). Set I consists of items of one facial expression with two emotion words. The participant has to select the word that matches the expression. Set II consists of items of one emotion word with two pictures of facial expressions. Here, the participant has to select the expression that matches the word. Set III contains items of two pictures of facial expressions and two emotions words. Again, the participant has to match the faces and the words. The outcome variable of the MENT is the number of errors. In previous studies where the validity of the MENT was assessed as a PVT for PTSD, 79% of suspected simulators of PTSD were detected and 96% of credible participants were assigned to the non-simulators (Morel & Shepherd, 2008).

#### Autism-Spectrum Quotient (AQ)

The AQ (Baron-Cohen et al., 2001) is a 50-item questionnaire developed to assess autistic characteristics in adults. Each item is a statement (e.g., "I enjoy doing things spontaneously") and the participant rates his/her degree of agreement (definitely agree, slightly agree, slightly disagree, and definitely disagree). Total score on the AQ can range from 0 to 50 and a score of 32 or higher in a general population sample indicates a clinical level of autistic traits (Woodbury-Smith et al., 2005). In the present study the Dutch version of the AQ was used, which has good internal consistency (α = .81), good test-retest reliability (α = .78), and good discriminative validity (Hoekstra et al., 2008).

#### Faux-Pas Test

The Dutch adult short version of the Faux-Pas Test (Spek et al., 2010; original version Stone et al., 1998) assesses

**Table 1.** Participants' background variables in the experimental and control condition of Study 1.

| | Simulation ($n = 26$) | | Control ($n = 26$) | |
|---|---|---|---|---|
| | *M (SD)* | Range | *M (SD)* | Range |
| Sex (males/females) | 5/21 | | 8/18 | |
| Age (in years) | 22.5[a] (7.3) | 18.3–46.5 | 21.0[a] (2.0) | 18.8–27.1 |
| Depressive symptoms (BDI-II) | 6.4 (4.5) | 0–18 | 7.9 (7.6) | 0–33 |
| Autism traits (AQ) | 9.5 (4.6) | 2-19 | 12.3 (5.8) | 4–25 |

[a]Based on $n = 23$ as not all participants responded when we asked for information regarding their date of birth.

BDI-II: Beck Depression Inventory-II; AQ: Autism-Spectrum Quotient.

theory of mind, the ability to understand mental states (Premack & Woodruff, 1978). It has been widely used in autism research and comprises nine stories of social situations. In five stories, a faux pas exists (a violation of social rules due to wrongful interpretation or an inability to perceive signals). The participant has to identify whether a faux pas occurred and answers faux pas related questions. Two control questions are asked to check correct understanding of the story. The outcome measure is the total number of correctly answered questions, with a maximum score of 56.

### Additional tests

Other tests, besides the MENT and the Faux-Pas Test, that were solely included in the (neuro)psychological test battery as fillers were (1) the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1964), which assesses verbal episodic memory (memorizing a series of non-related words); and (2) Visual Reproduction (a subtest from the Wechsler Memory Scale-III; Wechsler, 1997), which assesses visual episodic memory (memorizing a series of geometrical figures), please see the Supplementary Material (Appendix A) for a more elaborate description. As these tasks were considered filler tasks to convince the participants that it was a complete psychological assessment, the responses were neither coded or analyzed.

### Exit questionnaire

An exit questionnaire, delivered at the end of the psychological assessment, was used to assess if and how participants in the simulation condition used a strategy to simulate autism. For each measure, participants were asked whether they deliberately performed worse than they normally would have, performed better than they normally would have, or did not use a strategy to simulate autism. Furthermore, participants in the simulation condition were asked to rate on a 5-point Likert scale how motivated they had been to simulate autism (1= not motivated at all; 5 = very motivated) and how successful they thought they had been in simulating (1= not successful at all; 5 = very successful). Participants in the control condition were asked to rate how an individual with autism (with comparable age and IQ) would perform on each of the tests compared to their own performance. They could select: 1) worse than I, 2) better than I, or 3) similar as I. Also, they were asked to rate on a 5-point Likert scale how motivated they had been to do their best (1= not motivated at all; 5 = very motivated).

### Procedure

The project was approved by the appropriate institutional research ethics committee in 2014 [2014-BC-3951] and was in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All participants gave their written informed consent to participate in the study. Prior to the experiment, all participants received the AQ and an information brochure about autism traits in adults via email. The information brochure described social and communication impairments, emotion recognition problems, and restricted or repetitive interests and behaviors. Participants were instructed to fill in the AQ and read the digital brochure before they attended the experiment. Each participant was invited for an individual test session in the university lab and the session lasted approximately 80 minutes. The MENT was integrated in a test battery (see Supplementary Material), because we aimed to mimic a complete psychological assessment to increase similarity with a real life diagnostic test situation and to make it more challenging for the participants in the simulation condition. The order of the test battery varied and was counter-balanced across conditions in an even, nonrandom fashion. The MENT was either given early (after 15 minutes of testing) or late (after 60 minutes of testing) in the session.

Before the start of the test session, each participant was randomly assigned to the simulation or the control condition. The experimenter checked whether the participant had read the information brochure (see the Supplementary Material [Appendix B] for a brief section of the information brochure regarding emotion recognition) and asked the participant to mention a couple of autism characteristics. If the participant spontaneously mentioned "emotion recognition problems", the experimenter would continue with the experiment. If not spontaneously mentioned, the participant was prompted by the experimenter to tell more about the "social, communicative, and emotional problems" which autistic people are believed to experience. If the participant still did not bring up emotion recognition problems, the experimenter would mention these before continuing with the experiment.

Participants in the control condition received the following instruction from the experimenter: "I will administer several tests that a psychologist could use in diagnostic practice. It is important that you do your best, so that your test performances will correctly reflect your abilities."

Participants in the simulation condition received a scenario in order to facilitate the simulation of autism. Scenarios are frequently used in simulation studies and it is generally more difficult to detect simulation when the simulation scenario is relatively subtle and in a non-forensic context (Merckelbach et al., 2009). Because we wanted to motivate participants to simulate autism as credibly as they could, we added that the "most credible simulator of autism" would be rewarded €20,- upon completion of the data collection of the study. In reality, a participant in the simulation condition was *randomly* selected to receive the reward after data collection completion. Detailed instructions for the simulators can be found in the Supplementary Material (Appendix C).

After participants finished the test battery, they filled in an exit questionnaire. Participants in the simulation condition were instructed to stop simulating and answer the questions according to their own beliefs. All participants received credits for their participation. Upon request, participants received a brief description of the group results of the study.

### Statistical analyses

Exclusion criteria for the statistical analyses were (1) an AQ score of 32 or higher at screening, indicating a clinical level of autistic traits, and (2) responses on the exit questionnaire of participants in the simulation condition indicating that they had not used a strategy to simulate autism on the MENT or had not been motivated to simulate autism (score of 1 or 2 on a 5-point scale).

Following the statistical recommendations by Binder et al. (2014), we chose a one-tailed independent t-test (or a one-tailed Mann–Whitney U test in case of a non-normal distribution) to check whether the participants in the simulation condition made more errors on the MENT compared to the control group. To determine the optimal cutoff point on the MENT for distinguishing simulating participants from non-simulating participants, we used a receiver operating characteristic (ROC) curve analysis (Hosmer et al., 2013). We performed two additional analyses to check whether participants in the simulation condition were successfully simulating autistic traits. First, we used a paired t-test to compare their AQ scores during simulation to their own AQ scores during screening. Second, with an independent t-test (or a one-tailed Mann–Whitney U test in case of a non-normal distribution) we examined whether participants in the simulation condition performed more poorly on the Faux-Pas Test compared to the control group. Please note that data on key variables (excluding personal information to ensure participants' anonymity) can be found via the Open Science Framework (https://osf.io/qckhe/).

**Table 2.** Responses on the exit questionnaire of Study 1.

| | Simulation | | Control | |
|---|---|---|---|---|
| | "I performed intentionally worse on this task" | | "An individual with autism would perform worse than I" | |
| | Yes | No | Yes | No |
| Morel Emotional Numbing Test | 26 (100%) | 0 (0%) | 25 (96%) | 1 (4%) |
| Rey Auditory Verbal Learning Test[a] | 4 (15%) | 22 (85%) | 4 (17%) | 20 (83%) |
| Visual Reproduction (WMS-III) | 6 (23%) | 20 (77%) | 3 (12%) | 23 (88%) |
| Faux-Pas Test[b] | 24 (92%) | 2 (8%) | 24 (96%) | 1 (4%) |

WMS-III: Wechsler Memory Scale-III. [a]Based on $n = 24$ in control condition; [b]Based on $n = 25$ in control condition.

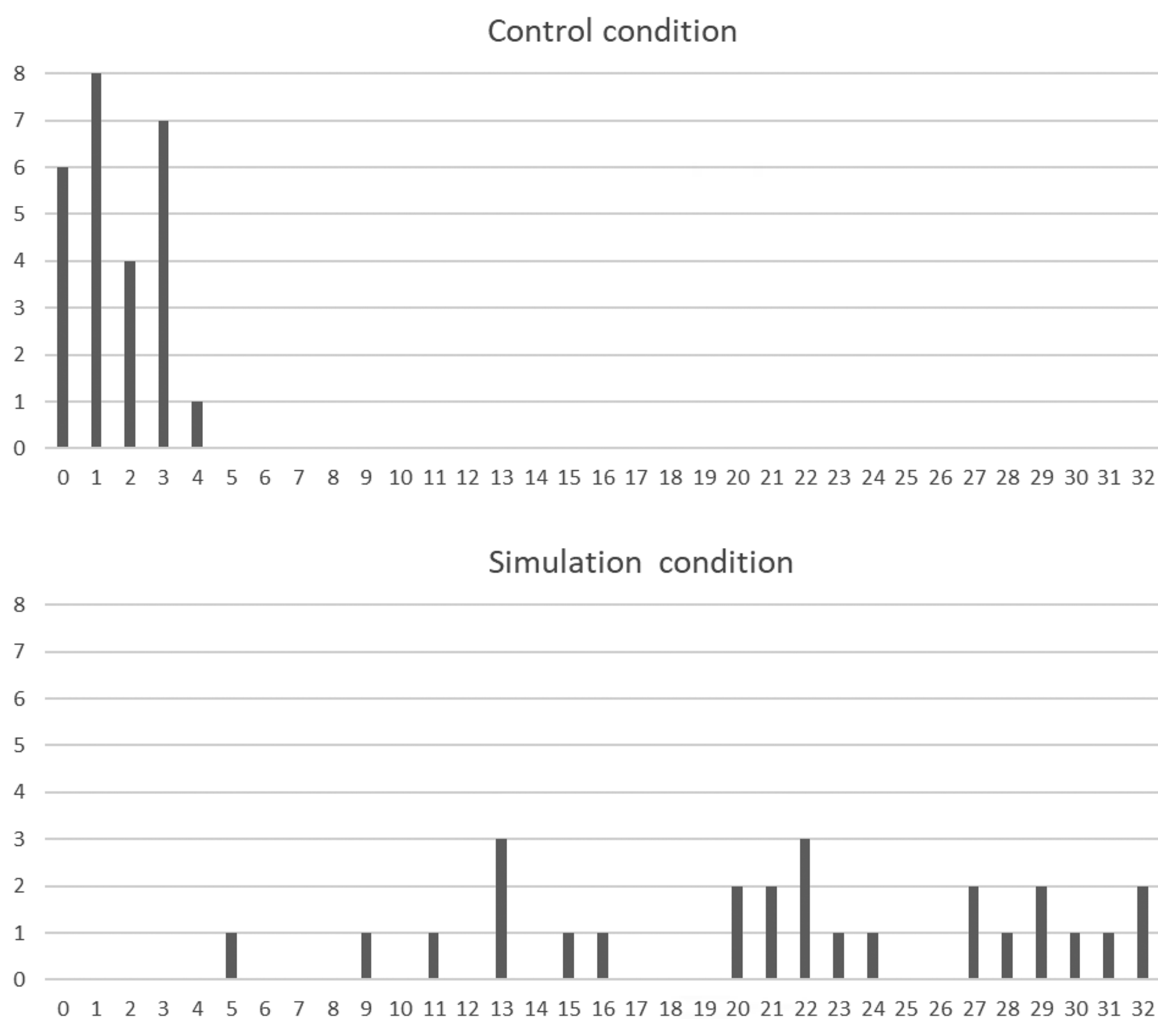



**Figure 1.** Distribution of number of errors (*x*-axis) on the MENT made by the number participants (*y*-axis) in the control (top panel) and simulation (bottom panel) condition.

## Study 1 results

### Exit interview data

Based on the exclusion criteria, we did not have to exclude any of the participants from the statistical analyses. First, looking at the data from the exit interview, all participants in the simulation condition reported that they had intentionally performed worse on the MENT than they normally would have (see Table 2). Furthermore, all but one (96%) participant in the control condition reported that they expected a person with autism to perform worse on the MENT than they had (see Table 2).

### MENT data

Because errors on the MENT were not normally distributed in the control condition (as can be seen in Figure 1) and log-transformation did not lead to a normal distribution, we used a Mann-Whitney U test to compare the error rate between the simulation and the control condition. As expected and in line with the exit interview data, participants instructed to simulate autism made significantly more errors on the MENT ($M = 21.3$, $SD = 7.6$) than those in the control condition ($M = 1.6$, $SD = 1.2$; $U = 0.00$, $Z = 6.2$, $p < .001$). The effect size was very large ($r = 0.9$, 95% CI [0.6, 1.1]).

The ROC curve analysis showed that a cutoff point of 4.5 errors on the MENT would be an optimal point to reach 100% specificity (every participant who was not simulating would be detected as a non-simulator) as well as 100% sensitivity (every participant who was simulating autism would be detected as a simulator). However, because the highest number of errors by a non-simulator (4) was very close to the lowest number of errors made by a simulator (5), and it is important to reduce the risk of falsely identifying a non-simulator as a simulator, we increased the cutoff point to 7 errors (increasing specificity at the cost of sensitivity; Sweet et al., 2021) based on the ROC curve. This new cutoff point led to 100% specificity and 96% sensitivity of the MENT.

### Simulation checks

AQ; When we compared AQ-scores of simulating participants to their own AQ-scores during screening, we found that AQ-scores during simulation were significantly increased ($t(25) = 33.6$, $p < .001$, Cohen's $d = 6.6$) and all individual AQ-scores during simulation were well above the clinical threshold of 32 ($M = 45.1$, $SD = 3.8$). This finding indicates that all participants in the simulation condition were able to simulate autism characteristics on the AQ.

Faux-Pas Test; In line with the responses given on the exit questionnaire, simulating participants performed significantly worse on the Faux-Pas Test ($M = 31.8$, $SD = 5.6$) compared to the participants in the control condition ($M = 50.7$, $SD = 4$; $U = 6$, $Z = 6.1$, $p < .001$). The effect size was large ($r = 0.8$, 95% CI [0.6, 1.1]).

## Discussion Study 1

Non-autistic students who were instructed to simulate autism made substantially more errors on the MENT compared to students who were instructed to do their best. Hence, instructed simulators could be reliably differentiated from non-simulators based on MENT performances. An analysis of the responses on the exit questionnaire revealed that all simulating participants had deliberately performed worse on the MENT and all but one participant in the control condition expected a person with autism to perform worse on the MENT. This indicates that participants seemed to believe that autistic people demonstrate poor facial emotion recognition, even when emotional displays are straightforward and simple. Together these findings support the idea that people perceive the MENT as a measure of autistic traits.

Participants instructed to simulate autism also reported clinical levels of autistic traits on the AQ and performed more poorly on the Faux-Pas Test compared to the control group. Previous studies have shown that simulating may be relatively easy when based on self-report questionnaires (Musso & Gouvier, 2014; Sweet et al., 2021; Tucha et al., 2009), and our AQ data support this. However, our Faux-Pas Test data suggest that classic tests to measure social cognition, a skill often said to be hampered in autistic individuals (Lever & Geurts, 2016), may be susceptible to simulation as well. This is particularly noteworthy as social cognition tests can be part of the autism diagnostic process. Simulation on more general cognitive tasks is also commonly observed in simulation studies focusing on cognitive profiles (e.g., Hardeman, 2006; Mestdagh et al., 2019).

Study 1 has some limitations that we should consider. First, we emphasized that credibility was important in the simulation condition and credibility would potentially be rewarded, yet the high AQ scores of the simulation group ($M = 45.1$), their poor performance on the Faux-Pas Test, and the high number of errors on the MENT suggest that some participants may have grossly exaggerated autistic traits rather than realistically portraying an autistic person. This is potentially problematic, because extreme underperformance by instructed simulators may not resemble performance by people who simulate or exaggerate in a clinical or forensic context. One should always keep in mind that simulation or exaggeration by instructed simulators is likely more easily detected by PVTs/SVTs than performances of those suspected of simulation or exaggeration in a clinical or forensic context (Sweet et al., 2021; Vickery et al., 2001). Moreover, the fact that all participants in Study 1 were first year psychology or psychobiology students increases the likelihood that they already had some basic knowledge of autism. Nonetheless, in future simulation studies it is recommendable to assess participants' knowledge about autism more systematically and also to actively involve autistic adults in the study so as to create better information about autism in adults and a more realistic simulation condition.

Note that participants made a deliberate decision to perform below their ability on some tests, but not the two memory tests, suggesting that they seemed at least partly successful in their understanding and display of an autistic

Table 3. Participants' background variables in the clinical and control group of Study 2.

| | Autism Group[a] ($n=25$) | | Control Group ($n=25$) | | Group comparison | | | |
|---|---|---|---|---|---|---|---|---|
| Measure | *M* (*SD*) | Range | *M* (*SD*) | Range | *t* value (48) | 95% CI | *p* | Cohen's *d* |
| Sex (males/females) | 15/10 | | 15/10 | | | | | |
| Age (years) | 48.9 (6.3) | 27–56 | 49 (6.5) | 27–57 | −0.11 | −3.84 to 3.44 | .91 | −0.02 |
| IQ | 124.8 (16.8) | 98–155 | 121.5 (13.9) | 100–149 | 0.75 | −5.49 to 12.05 | .46 | 0.22 |
| Autistic traits (AQ) | 30.9 (6.7) | 14–42 | 12 (5.4) | 5–25 | 11 | 15.43 to 22.33 | <.01 | 3.17 |

[a]Based on the Mini International Neuropsychiatric Interview eight autistic participants also met criteria for an anxiety disorder and five for ADHD.

pattern of performances. However, because the memory tests were primarily meant as filler tests, these data were not coded nor analyzed, which is an additional limitation of Study 1. Interestingly, the number of errors on the MENT by the instructed autism simulators approximated the average number of errors made by the group of veterans who were suspected of simulating PTSD (non-instructed simulators) in the Morel (1998) study. This tentatively suggests that people suspected of simulating may perform quite similarly as instructed simulators on the MENT, although further research in a clinical context is needed to confirm this.

Study 1 provides "proof of concept" of the MENT as a potential PVT for autism-related cognitive difficulties. Research with actually autistic adults is necessary to further test the validity of the MENT (Sweet et al., 2021). In Study 2, we therefore tested Hartman's third PVT criterion (insensitivity to the condition of interest) and second criterion (face-validity) by comparing MENT performances of adults with and without a clinical autism diagnosis. We hypothesized that autistic adults would perform similarly on the MENT as non-autistic adults.

## Study 2 method

### Participants

Twenty-six adults with a clinical autism diagnosis (autism group) and 26 non-autistic controls between 27 and 57 years and individually matched on age and biological sex (control group) were recruited for Study 2, which took place in 2014. Autistic participants were recruited via an adjacent, larger study at the University of Amsterdam with a focus on symptomatology, co-occurring conditions, and cognition in autistic adults (e.g., Geurts et al., 2021; Lever & Geurts, 2016). All participants in the autism group had received their clinical diagnosis in adulthood prior to (between 2006 and 2013) and independent from this study by a multidisciplinary team, including a psychiatrist or psychologist, according to DSM-IV-TR criteria (APA, 2000). Exclusion criteria were a current mood or psychotic disorder according to the Mini International Neuropsychiatric Interview (M.I.N.I., Van Vliet et al., 2000; original version Sheehan et al., 1998) and/or an estimated IQ below 80 as indicated by performance on two subtests (Vocabulary and Matrix Reasoning) of the Dutch version of the Wechsler Adult Intelligence Scale (WAIS-III, Uterwijk, 2000). One autistic participant had a current mood disorder and was therefore excluded from the analysis, along with their matched control. Control participants were recruited via the authors' social networks, matched on age and biological sex of the autistic participants. Exclusion criteria for adults in the control group were (1) close relatives with autism (i.e., parent, child, sibling) and/or (2) an AQ score above the cutoff of 32, indicating a relatively high level of autism traits (recommended cutoff for general population samples by Woodbury-Smith et al., 2005). Based on these criteria, none of the control participants had to be excluded from the analyses. Please note that six adults with a clinical autism diagnosis had an AQ score below the recommended cutoff for a referred clinical sample (<26; Woodbury-Smith et al., 2005). However, as they all had a clinical diagnosis and exploratory analyses revealed that results were similar with or without these participants, we chose not to use the AQ score as a stringent inclusion criterion.[1] Hence, we included 50 participants in our data analyses (see Table 3). IQ was matched on group level. Nobody in the control group used psychotropic medication on a permanent basis compared to eleven adults in the autism group. Five autistic adults used methylphenidate and abstained from its use 24 hours preceding and during their participation. Other medication was continued during participation as abstinence would influence daily functioning too much (3× antipsychotics, 6× antidepressants, 1× benzodiazepines).

### Measures

For a description of the MENT, the AQ, and the Faux-Pas Test we refer to the method section of Study 1. Because participants were informed that the aim of Study 2 was to examine emotion recognition in adults with and without autism, the Faux-Pas Test and the Emotion Recognition Task were included in the test battery. Other measures used to assess eligibility for the study are described in the Supplementary Material (Appendix A).

#### Emotion recognition task

The Emotion Recognition Test (ERT; Montagne et al., 2007) consists of short video clips of facial expressions (for details see Frigerio et al., 2002). The expressions on screen change from neutral to one of four different intensities (0–40%,

1. When the statistical analyses were repeated including only the participants that met all inclusion criteria ($n_1 = n_2 = 19$), the findings of the main confirmatory analysis were comparable with the current findings, although the $BF_{01}$ for the MENT and ERT slightly changed, hinting at undecisive evidence for similar MENT performances and moderate evidence for similar ERT performances.

0–60%, 0–80%, and 0–100%) of happiness, sadness, anger, fear, surprise, or disgust. Participants are instructed to select one of six emotion words to match the expression without time constraints. The outcome measure is the total number (range 0 to 96) of correctly identified expressions.

### Study purpose questionnaire

To evaluate whether the participants believed the study's supposed aim (examine emotion recognition) and whether they were blind to the assessment of performance validity, the experimenter asked participants: "What was, according to you, the purpose of this study?." Also, participants were asked to rate their ability to recognize emotions on a 100 millimeter visual analogue scale from '0' ("I am unable to recognize emotions") to "100" ("I am very good at recognizing emotions"). The outcome measure is the distance in millimeters on the scale.

### Procedure

The project was approved by the appropriate institutional research ethics committee in 2014 [2014-BC-3472] and was in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All participants were informed that the research aim was to determine emotion recognition abilities in autistic people. Once written informed consent was obtained, participants were tested in a 2-hour session, including a short break. In addition to the earlier described materials, six other questionnaires were administered for the purpose of the adjacent study. The task order varied and was counterbalanced across participants.

Upon study completion, all participants received full debriefing about the study. Autistic participants received a reimbursement of travel expenses (maximum of €20) and control participants were compensated with €15 for their participation.

### Data analyses

An a priori power analysis ($1 - \beta = .95$, $\alpha = .05$, effect size = .45) for typical frequentist analyses showed that a sample size of 45 participants was required to detect potential differences in means and frequencies between the autism and control group with a medium to large effect size (Cohen, 1992). However, given our hypothesis (i.e., the autism and control group will *not* differ in MENT performance) we used Bayesian hypothesis testing (Jeffreys, 1961). This statistical procedure calculates the extent to which the data support hypothesis A (e.g., null hypothesis) over hypothesis B (e.g., alternative hypothesis), expressed by a Bayes factor (BF; Rouder et al., 2009). We used Bayesian interference primarily to determine whether there is evidence in favor of the null hypothesis, so called "evidence of absence" (Dienes, 2014). For example, a $BF_{01}$ of 4 indicates that the data are four times more likely to have occurred under the null hypothesis (0) than under the alternative hypothesis (1), a $BF_{10}$ of 4 indicates the opposite. Even though BFs can be interpreted continuously, a rule of thumb which is often used to categorize evidential strength (see van Doorn et al., 2021) is: BFs between 0.3 and 3 are defined as undecisive (i.e., not reflecting enough evidence for either hypothesis), BFs between 3 and 10 indicate moderate evidence, between 10 and 30 strong evidence, between 30 and 100 very strong evidence, and above 100 extreme evidence for the null or the alternative hypothesis. An important aspect of Bayesian statistics is that one needs to define the priors (i.e., belief about the effect before observing the data). In this case we used the default setting in JASP which is a Cauchy prior set at 0.707. To assess the robustness of the BFs obtained we included a robustness check using a range of different prior distributions (BF robustness plots are provided in Appendix D, Supplementary Material). If a BF did not seem robust, this is noted in the results section.

As an additional check for participants' blindness to the study's true purpose, answers on the study purpose questionnaire were qualitatively evaluated to verify whether participants were indeed blind to performance validity testing. Additionally, we ran an independent t-test on self-rated emotion recognition ability to test whether autistic adults rated themselves less able than adults without an autism diagnosis. Furthermore, to determine whether theory of mind and/or emotion recognition was poorer in the autism group compared to the control group, two Bayesian independent t-tests were performed with Faux-Pas Test/ERT performance as dependent variable(s).

BFs and robustness checks were computed using JASP 0.17.3 (JASP Team, 2023). All other analyses were done using SPSS V22 (IBM Corp, 2013). Assumptions for the mentioned analyses were checked, as well as outliers using the Outlier Labeling Rule (Hoaglin & Iglewicz, 1987). With this method, the bounds of the lower and upper quartile of the data are estimated and data points that exceed these bounds are identified as outliers. Please note that the data including key dependent measures (excluding personal information to ensure anonymity) can be found at OSF (https://osf.io/qckhe/) and main JASP results can be found in Appendix D (Supplementary Material).

Finally, the cutoff point as determined by the ROC curve analysis from Study 1 was applied to the non-simulating adults of Study 2 to check how many individuals were correctly identified as non-simulators. Furthermore, we did a one-sample t-test (or a Mann–Whitney U test in case of a non-normal distribution) to compare the mean number of errors on the MENT by the autistic adults in Study 2 to the mean number of errors made by the simulating group in Study 1.

## Study 2 results

### Assumptions and outliers

MENT performance was not normally distributed across groups. Within the autism group specifically, MENT performance was positively skewed. Therefore, data were transformed using log transformation (Field, 2009). After transformation, all assumptions were met to execute the Bayesian independent t-test on MENT performance (Field, 2009). No significant outliers were detected within the transformed MENT data.

Table 4. Descriptives and results of confirmatory and exploratory analyses of Study 2.

| | Autism group ($n=25$) | | Control Group ($n=25$) | | Group comparison | | | |
|---|---|---|---|---|---|---|---|---|
| Measure | *M* (*SD*) | Range | *M* (*SD*) | Range | *t* value (48) | 95% *CI* | *p* | Cohen's *d* |
| MENT (# of errors) | 1.9 (1.5) | 0–6 | 2.3 (1.9) | 0–8 | −0.42 | −0.16 to 0.11 | .67 | −0.12 |
| Rated emotion recognition (%) | 57.6 (19.7) | 20–85 | 79.4 (10.2) | 60–100 | −4.93 | −30.83 to −12.84 | <.01 | −1.42 |
| Faux-Pas Test | 45.8 (3.7) | 37–51 | 48.9 (4) | 43–55 | −2.89 | −5.36 to −0.96 | .01 | −0.82 |
| ERT (total score) | 60.9 (9.4) | 42–79 | 58.9 (8.3) | 42–74 | 0.77 | −3.12 to 6.96 | .45 | 0.23 |

*MENT* = Morel Emotional Numbing Test, *ERT* = Emotion Recognition Test.

With regard to all exploratory analyses, homogeneity of variance was not met for the data of self-rated emotion recognition ability thus the test values which take this violation into account are reported. All further assumptions were met to execute the exploratory analyses, and no outliers were detected.

### Matching

Individual matching on age and biological sex was successful. In addition, IQ was successfully matched on group level. As expected, the autism group had significantly higher AQ scores than the control group (see Table 3).

### Confirmatory analyses: MENT

In line with our hypothesis, autistic participants performed equally well on the log transformed MENT as their matched controls as $BF_{01}$ = 3.3 (see Table 4 and Appendix D, Supplementary Material). The observed data are, therefore, 3.3 times more likely to have occurred under the null hypothesis than under the alternative hypothesis. Thus, there is moderate evidence in favor of the null hypothesis (i.e., no difference between groups) in comparison to the alternative hypothesis. The error percentage is <0.01 which indicates that the obtained BF is sufficiently stable.

### Exploratory analyses: study purpose questionnaire, Faux-Pas Test, and ERT

#### Purpose of the study questionnaire

None of the participants indicated any awareness that validity of their performance was evaluated in the current study. All reported that the purpose of the study was to investigate emotion recognition in autistic individuals and controls. One autistic participant stated that the purpose of the study was to exclude poor emotion recognition as a key characteristic of autism. Autistic participants rated their emotion recognition ability significantly lower compared to the control group (see Table 4), with $BF_{10}$=1600. The observed data are, therefore, 1600 times more likely to have occurred under the alternative hypothesis than under the null hypothesis. Thus, there is extreme evidence in favor of the alternative hypothesis (i.e., the two groups differ in their ratings) in comparison to the null hypothesis. The error percentage is <0.001 which indicates that the obtained BF is stable.

#### Faux-Pas Test

Participants with a clinical autism diagnosis performed significantly worse on the Faux-Pas Test in comparison to their matched controls (see Table 4), $BF_{10}$=7.5. Given the data, there is moderate evidence in favor of the alternative hypothesis in comparison to the null hypothesis. The error percentage is <0.001, indicating that the obtained BF is stable.

#### ERT

Participants with a clinical autism did not perform significantly worse on the ERT in comparison to their matched controls (see Table 4), $BF_{01}$ = 2.8. The observed data are therefore 2.8 times more likely to have occurred under the null hypothesis than under the alternative hypothesis, indicating that the evidence is undecisive. The error percentage is <0.01 which indicates that the obtained BF is sufficiently stable.

### Exploratory comparison with Study 1

We applied a cutoff score of 7 errors on the MENT data (based on Study 1) to check whether all non-simulators in Study 2—with and without an autism diagnosis—would be correctly identified as a non-simulator. All but one participant in Study 2 ($n=49$) were identified as a non-simulator, indicating a specificity of 98%. A comparison between the autistic adults in Study 2 and the participants in the simulation condition of Study 1 showed that simulating participants made significantly more errors on the MENT compared to the autistic adults ($U=1.5$, $Z$ =-6.1, $p$ < .001). The effect size was very large ($r=-0.9$, 95% CI [0.6, 1.1]).

## Discussion Study 2

As hypothesized, adults with and without an autism diagnosis did not differ in their MENT performances. In fact, the group of autistic adults performed substantially better on the MENT (made far fewer mistakes) than the group of instructed simulators in Study 1. This is the first tentative evidence that the MENT is insensitive to a clinical diagnosis of autism in adults with a normal to high IQ and thus seems to fulfill the third Hartman (2002) criterion of an effective PVT. Furthermore, exploratory analyses demonstrated that autistic adults generally believed that their emotion recognition was poor, satisfying criterion 2 (i.e., people have the false belief that the MENT assesses autistic traits).

Hence, the MENT appears to be a face-valid measure of autistic traits, even though autism is in fact not related to performance on the MENT. The outcomes of Study 2 offer additional support for the MENT as a potential PVT for autism-related cognitive difficulties.

Alternative explanations and limitations of Study 2 must also be considered. First, while we showed that the MENT is insensitive to an autism diagnosis, it is uncertain to what extent our results generalize to autistic individuals of all ages and (below) average intellectual abilities. Average to high intellectual abilities are often observed in autistic people who are diagnosed in (late) adulthood (Lai and Baron-Cohen (2015). These late-diagnosed people may have been missed earlier because their strong cognitive skills in part compensate for social impairments. Hence, one might argue that the autistic adults in Study 2 may have performed exceptionally well on the MENT due to their high IQ levels. However, research has shown that, despite general emotion recognition problems in autism (Uljarevic & Hamilton, 2013), most autistic individuals with an average to high IQ are in fact well able to recognize the type of emotions we included (Nuske et al., 2013). Also, their slightly, but significantly worse performance on the Faux-Pas Test (compared to the non-autistic group) demonstrates differences in grasping others' intentions and eliminates the alternative interpretation that their IQ levels conceal any autism-related challenges. Second, and closely related to the previous point, our participants with an autism diagnosis were all diagnosed as an adult and this possibly restricts the generalization of our findings to adults with an earlier (child) autism diagnosis. Yet, we believe the late age of diagnosis should in fact be viewed as a strength of the study, because the MENT may ultimately—if deemed valid—be used as a PVT in psychological assessments of autism in adults (rather than adults already diagnosed in childhood). Third, the relatively high proportion of autistic women in our clinical sample (compared to the prevalence of women in the autistic population) may have led to a better overall performance on the MENT, because females tend to outperform males on emotion recognition tasks (Montagne et al., 2005). However, males and females did not differ in their performance on the MENT and a rerun of our (underpowered) analyses without females supported the observed null finding.[2] Moreover, given the rise of adult autism diagnoses in females in particular (Huang et al., 2021, Underwood et al., 2022; but see Rutherford et al., 2016), the high proportion of autistic females in our study may also be considered a strength.

## General discussion

Exaggeration and simulation of autistic traits or autism-related cognitive challenges have hardly been examined thus far, likely because these are highly controversial topics. However, in this paper we argue it is possible that autistic traits can be simulated or exaggerated by some adults based on exaggeration and simulation rates of other conditions such as ADHD (Sullivan et al., 2007; Tucha et al., 2009), clinicians' suspicion of feigning autism in adults (Murphy & Broyd, 2023), and the less than perfect practice of diagnosing autism in adulthood (Bal & Taylor, 2019), including a strong reliance on self-report. Indeed, in Study 1, we found that autism simulators scored well above the clinical threshold on the Autism-Spectrum Quotient, a widely used screener questionnaire for autistic traits in adults (Hoekstra et al., 2008). This seems to suggest that simulators can report behavioral tendencies and preferences corresponding with an autistic profile. This warrants caution using the AQ in the autism diagnostic process and further confirms the need for PVTs/SVTs. We further argue that simulation or exaggeration of autistic traits or autism-related cognitive difficulties may not only be possible, but may also be beneficial for some people (e.g., reduced stigma compared to other psychiatric diagnoses (Gillespie-Lynch et al., 2021), financial gains for research participation (Pellicano et al., 2023)). To reduce the likelihood of false positives, it is crucial to develop validated PVTs for autism-related difficulties. PVTs have the advantage over SVTs of relying on performance rather than self-reported traits. Indeed, a recent factor analysis study demonstrated that SVTs, rather than PVTs, loaded onto the same latent factor as symptom reports (Ord et al., 2021). Additionally, the same study showed that PVTs and SVTs assess distinct, but related constructs. Therefore, failing a PVT for autism-related cognitive difficulties is likely associated with a non-credible presentation of autism symptoms, but this remains to be tested in clinical practice. Only when we have a validated PVT we can empirically test whether exaggeration/simulation is a problem in some adults with possible autism.

The current two studies were the first to consider the MENT as a potential PVT for autism-related cognitive difficulties. According to Hartman's criteria (Hartman, 2002), a PVT should encourage a simulator to perform beneath their true level of ability (criterion 1), should appear to measure traits of the condition of interest (criterion 2), and should be insensitive to the condition of interest (criterion 3). The MENT appears to fulfill all of these criteria in this first study of adults instructed to simulate autism and adults with a clinical diagnosis of autism. First, and crucially, instructed simulators could be reliably differentiated from non-simulators based on MENT performances (satisfying criterion 1). Second, participants in Studies 1 and 2 believed poor emotion recognition or poor performance on the MENT would characterize autistic people (satisfying criterion 2). Third, adults diagnosed with autism in adulthood in fact performed equally well on the MENT as adults without autism (satisfying criterion 3). When we applied a

2. We performed an exploratory t-test with biological sex as the independent variable and transformed MENT scores as the dependent variable. The mean error score on the MENT ($M_{\text{males}}$ = 1.9, $SD_{\text{males}}$ = 1.5 and $M_{\text{females}}$ = 2.4, $SD_{\text{females}}$ = 1.9) did not differ across biological sex ($t(48) = -0.67$, $p = .51$). The rerun of our confirmatory analysis without females showed undecisive evidence for similar MENT performances of the autism ($M = 1.6$, SD = 0.9) and control group ($M = 2.2$, SD = 2.0; $t(28) = -0.65$, $p = .52$, $BF_{01} = 2.5$).

relatively conservative cutoff point of seven errors, we reached a specificity of ≥98% (identifying 100% non-simulators as non-simulator in Study 1 and 98% in Study 2) and a sensitivity of 96% (identifying 96% of simulators as simulator). Summing up, our findings provide first empirical evidence for the validity of the MENT as a potential PVT for autism-related cognitive difficulties.

Individual PVT results should be interpreted carefully. Because feigning can best be understood as a dimensional construct (Merten & Rogers, 2017), any cutoff score on a PVT creates an artificial categorical distinction. Therefore, examiners should be aware that performances close to the recommended cutoff may either be genuine or exaggerated/simulated. Furthermore, non-credible performance on a single PVT should not be taken as a definite indicator of feigning or exaggerating (Sweet et al., 2021). Instead, someone's complete test results and background should be taken into consideration, including the results of multiple PVTs and a person's cultural and ethnical background (Sweet et al., 2021), as the validity of PVTs and response styles may vary across culturally diverse groups (Merten & Rogers, 2017).

Autism-specific PVTs are relevant for actually autistic individuals, clinicians, and researchers. A successful method to detect autism-related exaggeration/simulation will likely (1) optimize care and support for individuals with a verified autism diagnosis, due to reduced "noise" in the autism group, and (2) reduce misconceptions of autism (Fombonne, 2023; Tucha et al., 2009). While a PVT may be particularly valuable in cases where the external benefits of an adult autism diagnosis are relatively high, such as forensic cases (e.g., in case of simulation, Maras et al., 2019), it is also important when a clinician wants to map someone's psychological strengths and weaknesses (e.g., in case of exaggeration, Bush et al., 2005). Finally, it is worth mentioning that people performing poorly on the MENT in a clinical context may still require clinical care and support as exaggeration/simulation may be a sign of mental health problems and unmet needs. Indeed, simulation and exaggeration of particular psychological/cognitive problems can co-exist with veridical psychological/cognitive problems (Merten & Merckelbach, 2013).

For now, our studies regarding the usefulness of the MENT as a PVT for autism-related cognitive difficulties provide evidence in favor of the validity of this measure. However, given the artificial lab conditions of our studies (e.g., simulation with an emphasis on emotion recognition problems), research in a clinical setting is necessary to further replicate or nuance our findings. It is likely that simulators in a real-world context show less extreme underperformance on the MENT. The next step therefore is to test the usability of the MENT to determine how people referred with a suspicion of autism perform on the MENT as compared to simulators and already diagnosed autistic adults.

## Acknowledgment

We thank both the participants and the mental health care institutions (Dr. Leo Kannerhuis, Autisme Team Noord Holland, Centrum Autisme GGZ Noord Holland Noord, GGZ Breburg) for their efforts. We would like to especially thank the participants and the students who helped with the data collection.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by Nederlandse Organisatie voor Wetenschappelijk Onderzoek.

## ORCID

Anke M. Scheeren http://orcid.org/0000-0001-7530-3354

## References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental Disorders.* (4th ed. text revision).

Bal, V. H., & Taylor, J. L. (2019). Advancing understanding of adults: The role of diagnostic confirmation and sample description. *Autism: The International Journal of Research and Practice*, *23*(4), 807–810. https://doi.org/10.1177/1362361319847547

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:100565341147

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for beck depression inventory-II*. Psychological Corporation.

Binder, L. M., Larrabee, G. J., & Millis, S. R. (2014). Intent to fail: Significance testing of forced choice test results. *The Clinical Neuropsychologist*, *28*(8), 1366–1375. https://doi.org/10.1080/13854046.2014.978383

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity—NAN policy & planning committee. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *20*(4), 419–426. https://doi.org/10.1016/j.acn.2005.02.002

Cage, E., Di Monaco, J., & Newell, V. (2019). Understanding, attitudes and dehumanisation towards autistic people. *Autism: The International Journal of Research and Practice*, *23*(6), 1373–1383. https://doi.org/10.1177/1362361318811290

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Dandachi-FitzGerald, B., Ponds, R. W. H. M., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of Neuropsychologists in six European countries. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *28*(8), 771–783. https://doi.org/10.1093/arclin/act073

Dandachi-FitzGerald, B., Ponds, R. W. H. M., Peters, M. J. V., & Merckelbach, H. (2011). Cognitive underperformance and symptom over-reporting in a mixed psychiatric sample. *The Clinical Neuropsychologist*, *25*(5), 812–828. https://doi.org/10.1080/13854046.2011.583280

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Feldman, D. B., & Crandall, C. S. (2007). Dimensions of mental illness stigma: What about mental illness causes social rejection? *Journal of Social and Clinical Psychology*, *26*(2), 137–154. https://doi.org/10.1521/jscp.2007.26.2.137
Field, A. P. (2009). *Discovering statistics using SPSS* (3rd ed.). SAGE Publications.
Fombonne, E. (2023). Editorial: Is autism overdiagnosed? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *64*(5), 711–714. https://doi.org/10.1111/jcpp.13806
Frigerio, E., Burt, D. M., Montagne, B., Murray, L. K., & Perrett, D. I. (2002). Facial affect perception in alcoholics. *Psychiatry Research*, *113*(1–2), 161–171. https://doi.org/10.1016/S0165-1781(02)00244-5
Geraerts, E., Merckelbach, H., & Jelicic, M. (2007). Het simuleren van posttraumatische stressymptomen: de Nederlandse versie van de Morel Emotional Numbing Test (MENT). *Neuropraxis*, *11*(1), 8–12. https://doi.org/10.1007/BF03079115
Geurts, H. M., Agelink van Rentergem, J. A. A., Radhoe, T., Torenvliet, C., Van der Putten, W. J., & Groenman, A. P. (2021). Ageing and heterogeneity regarding autism spectrum conditions: A protocol paper of an accelerated longitudinal study. *BMJ Open*, *11*(3), e040943. https://doi.org/10.1136/bmjopen-2020-040943
Gillespie-Lynch, K., Bublitz, D., Donachie, A., Wong, V., Brooks, P. J., & D'Onofrio, J. (2017). "For a long time our voices have been hushed": Using student perspectives to develop supports for neurodiverse college students. *Frontiers in Psychology*, *8*, 544. https://doi.org/10.3389/fpsyg.2017.00544
Gillespie-Lynch, K., Daou, N., Obeid, R., Reardon, S., Khan, S., & Goldknopf, E. J. (2021). What contributes to stigma towards autistic university students and students with other diagnoses? *Journal of Autism and Developmental Disorders*, *51*(2), 459–475. https://doi.org/10.1007/s10803-020-04556-7
Hardeman, C. (2006). Malingeren in de neuropsychologie: ontwikkeling en interpretatie van tests voor het detecteren van een suboptimale inzet. *Neuropraxis*, *10*(4), 102–109. https://doi.org/10.1007/BF03079092
Hartman, D. E. (2002). The unexamined lie is a lie worth fibbing. Neuropsychological malingering and the Word Memory Test. *Archives of Clinical Neuropsychology*, *17*(7), 709–714. https://doi.org/10.1016/S0887-6177(01)00172-X
Hoaglin, M. F. D. C., & Iglewicz, B. (1987). Fine tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, *82*(400), 1147–1149. https://doi.org/10.1080/01621459.1987.10478551
Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders*, *38*(8), 1555–1566. https://doi.org/10.1007/s10803-008-0538-x
Hoelzle, J. B., Ritchie, K. A., Marshall, P. S., Vogt, E. M., & Marra, D. E. (2019). Erroneous conclusions: The impact of failing to identify invalid symptom presentation when conducting adult attention-deficit/hyperactivity disorder (ADHD) research. *Psychological Assessment*, *31*(9), 1174–1179. https://doi.org/10.1037/pas0000752
Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (Eds.). (2013). *Assessing the fit of the model (ch. 5)*. In *Applied logisitic regression* (3 ed., pp. 153–225). John Wiley & Sons.
Huang, Y., Arnold, S. R. C., Foley, K.-R., Lawson, L. P., Richdale, A. L., & Trollor, J. N. (2021). Factors associated with age at autism diagnosis in a community sample of Australian adults. *Autism Research: official Journal of the International Society for Autism Research*, *14*(12), 2677–2687. https://doi.org/10.1002/aur.2610
IBM Corp. (2013). *IBM SPSS statistics for windows (Version 22.0).*
JASP Team. (2023). JASP (Version 0.17.3)[Computer software].
Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.
Lai, M.-C., & Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. *The Lancet. Psychiatry*, *2*(11), 1013–1027. https://doi.org/10.1016/S2215-0366(15)00277-1
Lever, A. G., & Geurts, H. M. (2016). Age-related differences in cognition across the adult lifespan in autism spectrum disorder. *Autism Research: official Journal of the International Society for Autism Research*, *9*(6), 666–676. https://doi.org/10.1002/aur.1545
Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223. https://doi.org/10.1023/A:1005592401947
Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule* (2nd ed.). Western Psychological Services.
Maras, K., Marshall, I., & Sands, C. (2019). Mock Juror perceptions of credibility and culpability in an autistic defendant. *Journal of Autism and Developmental Disorders*, *49*(3), 996–1010. https://doi.org/10.1007/s10803-018-3803-7
Merckelbach, H., Smeets, T., & Jelicic, M. (2009). Experimental simulation: Type of malingering scenario makes a difference. *Journal of Forensic Psychiatry & Psychology*, *20*(3), 378–386. https://doi.org/10.1080/14789940802456686
Merten, T., & Merckelbach, H. (2013). Introduction to malingering research and symptom validity assessment. *Journal of Experimental Psychopathology*, *4*(1), 3–5. https://doi.org/10.1177/204380871300400102
Merten, T., & Rogers, R. (2017). An international perspective on feigned mental disabilities: Conceptual issues and continuing controversies. *Behavioral Sciences & the Law*, *35*(2), 97–112. https://doi.org/10.1002/bsl.2274
Mestdagh, I., van Bergen, L., Kocken, C., Heyvaert, V., Cras, P., & Van Den Eede, F. (2019). Diagnosing solvent-induced chronic toxic encephalopathy: The effect of underperformance in neuropsychological testing. *International Journal of Psychiatry in Clinical Practice*, *23*(3), 171–177. https://doi.org/10.1080/13651501.2019.1571210
Monaro, M., Toncini, A., Ferracuti, S., Tessari, G., Vaccaro, M. G., De Fazio, P., Pigato, G., Meneghel, T., Scarpazza, C., & Sartori, G. (2018). The detection of malingering: A new tool to identify made-up depression. *Frontiers in Psychiatry*, *9*, 249. https://doi.org/10.3389/fpsyt.2018.00249
Montagne, B., Kessels, R. C., De Haan, E. H. F., & Perrett, D. I. (2007). The emotion recognition task: A paradigm to measure the perception of facial emotional expressions at different intensities. *Perceptual and Motor Skills*, *104*(2), 589–598. https://doi.org/10.2466/pms.104.2.589-598
Montagne, B., Kessels, R. P. C., Frigerio, E., de Haan, E. H. F., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? *Cognitive Processing*, *6*(2), 136–141. https://doi.org/10.1007/s10339-005-0050-6
Morel, K. (1998). Development and preliminary validation of a forced-choice test of response bias for posttraumatic stress disorder. *Journal of Personality Assessment*, *70*(2), 299–314. https://doi.org/10.1207/s15327752jpa7002_8
Morel, K. R., & Shepherd, B. E. (2008). Meta-analysis of the morel emotional numbing test for PTSD: Comment on Singh, Avasthi, & Grover. *German Journal of Psychiatry*, *11*(1), 128–131.
Murphy, D., & Broyd, J. G. (2023). Suspected feigning of autism in adults: a clinician survey, indications and proposed guidelines. *Advances in Autism*, *9*(1), 29–41. https://doi.org/10.1108/AIA-11-2021-0044
Musso, M. W., & Gouvier, W. D. (2014). "Why is this so hard?" A review of detection of malingered ADHD in college students. *Journal of Attention Disorders*, *18*(3), 186–201. https://doi.org/10.1177/1087054712441970
National Institute for Health and Clinical Excellence (NICE). (2012). *Autism: recognition, referral, diagnosis and management of adults on the autism spectrum*. National clinical guideline no 142. National Institute for Health and Clinical Excellence.
Nuske, H. J., Vivanti, G., & Dissanayake, C. (2013). Are emotion impairments unique to, universal, or specific in autism spectrum disorder? A comprehensive review. *Cognition & Emotion*, *27*(6), 1042–1061. https://doi.org/10.1080/02699931.2012.762900
Nylander, L., Holmqvist, M., Gustafson, L., & Gillberg, C. (2013). Attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASC) in adult psychiatry. A 20-year register study. *Nordic Journal of Psychiatry*, *67*(5), 344–350. https://doi.org/10.3109/08039488.2012.748824

Ord, A. S., Shura, R. D., Sansone, A. R., Martindale, S. L., Taber, K. H., & Rowland, J. A. (2021). Performance validity and symptom validity tests: Are they measuring different constructs? *Neuropsychology*, *35*(3), 241–251. https://doi.org/10.1037/neu0000722

Pellicano, E., Adams, D., Crane, L., Hollingue, C., Allen, C., Almendinger, K., Botha, M., Haar, T., Kapp, S. K., & Wheeley, E. (2023). Letter to the editor: A possible threat to data integrity for online qualitative autism research. *Autism: The International Journal of Research and Practice*, *0*(0), 13623613231174543. https://doi.org/10.1177/13623613231174543

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Rey, A. (1964). *L'examen clinique en psychologie*. Presses Universitaires de France.

Rouder, J. N., Speckman, L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rutherford, M., McKenzie, K., Johnson, T., Catchpole, C., O'Hare, A., McClure, I., Forsyth, K., McCartney, D., & Murray, A. (2016). Gender ratio in a clinical population sample, age of diagnosis and duration of assessment in children and adults with autism spectrum disorder. *Autism: The International Journal of Research and Practice*, *20*(5), 628–634. https://doi.org/10.1177/1362361315617879

Sheehan, D. V., Lecrubier, Y., Harnett-Sheehan, K. H., Amorim, P., Janavs, J., & Weiler, E. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, *59*(Suppl 20), 22–33.

Shulman, C., Rice, C. E., Morrier, M. J., & Esler, A. (2020). The role of diagnostic instruments in dual and differential diagnosis in autism spectrum disorder across the lifespan. *Child and Adolescent Psychiatric Clinics of North America*, *29*(2), 275–299. https://doi.org/10.1016/j.chc.2020.01.002

Spek, A. A., Scholte, E. M., & Van Berckelaer-Onnes, I. A. (2010). Theory of mind in adults with HFA and asperger syndrome. *Journal of Autism and Developmental Disorders*, *40*(3), 280–289. https://doi.org/10.1007/s10803-009-0860-y

Stone, V., Baron-Cohen, S., & Knight, R. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*(5), 640–656. https://doi.org/10.1162/089892998562942

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology*, *14*(3), 189–207. https://doi.org/10.1080/09084280701509083

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., & Suhr, J. A, Conference, P. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *35*(6), 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

Taylor, L. J., Eapen, V., Maybery, M. T., Midford, S., Paynter, J., Quarmby, L., Smith, T., Williams, K., & Whitehouse, A. J. O. (2016). Diagnostic evaluation for autism spectrum disorder: A survey of health professionals in Australia. *BMJ Open*, *6*(9), e012517. https://doi.org/10.1136/bmjopen-2016-012517

Timimi, S., Milton, D., Bovell, V., Kapp, S., & Russell, G. (2019). Deconstructing diagnosis: Four commentaries on a diagnostic tool to assess individuals for autism spectrum disorders. *Autonomy (Birmingham, England)*, *1*(6), AR26.

Tucha, L., Sontag, T. A., Walitza, S., & Lange, K. W. (2009). Detection of malingered attention deficit hyperactivity disorder. *Attention Deficit and Hyperactivity Disorders*, *1*(1), 47–53. https://doi.org/10.1007/s12402-009-0007-1

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders*, *43*(7), 1517–1526. https://doi.org/10.1007/s10803-012-1695-5

Underwood, J. F., DelPozo-Banos, M., Frizzati, A., John, A., & Hall, J. (2022). Evidence of increasing recorded diagnosis of autism spectrum disorders in Wales, UK: An e-cohort study. *Autism: The International Journal of Research and Practice*, *26*(6), 1499–1508. https://doi.org/10.1177/13623613211059674

Uterwijk, J. (2000). *WAIS-III. Nederlandstalige Bewerking. Technische Handleiding*. Swets & Zeitlinger.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*(3), 813–826. https://doi.org/10.3758/s13423-020-01798-5

Van Vliet, I. M., Leroy, H., & Van Megen, H. J. G. M. (2000). *M.I.N.I. Plus: M.I.N.I. internationaal neuropsychiatrisch interview: Nederlandse versie 5.0.0*. http://www.tijdschriftvoorpsychiatrie.nl/assets/measuringinstruments/meetinstrumenten_90pdf.pdf.

Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, *16*(1), 45–73. https://doi.org/10.1016/S0887-6177(99)00058-X

Wechsler, D. (1997). *Wechsler Memory Scale* (3rd ed). The Psychological Corporation.

Wigham, S., Rodgers, J., Berney, T., Le Couteur, A., Ingham, B., & Parr, J. R. (2019). Psychometric properties of questionnaires and diagnostic measures for autism spectrum disorders in adults: A systematic review. *Autism: The International Journal of Research and Practice*, *23*(2), 287–305. https://doi.org/10.1177/1362361317748245

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*, *35*(3), 331–335. https://doi.org/10.1007/s10803-005-3300-7