Autism in Aotearoa: Is the RAADS-14 a valid tool for a New Zealand population

Ms Sarah M. Kember

Dr Matt N. Williams

Massey University

This version of the article is an accepted manuscript (postprint), and may not completely replicate the final version of record published in the *European Journal of Psychological Assessment* at https://doi.org/10.1027/1015-5759/a000598.

Author Note

Ms Sarah M. Kember: https://orcid.org/0000-0002-0352-3283

Dr Matt N. Williams: https://orcid.org/0000-0002-0571-215X

Ms Kember is a postgraduate student of psychology at Massey University. Dr Williams is a lecturer in this department and has supervised this project. The authors have no known conflict of interest to declare. Correspondence concerning this article may be addressed to Ms Kember by email to sarah.kember.3@uni.massey.ac.nz. The authors extend their thanks to Autism New Zealand and Altogether Autism for their assistance, and to all the participants who so generously gave their time to this research.

Table of Contents

Abstract

Screening measures for autism spectrum disorder (ASD) are important tools for clinicians and researchers. However, where a measure developed and validated for one population is used with another, its performance in this new context must be carefully examined. The RAADS-14, a brief ASD screen developed in Sweden, was evaluated with a sample of New Zealand adults (N = 387), 41 of whom self-reported a prior diagnosis of ASD. The convergent validity of the RAADS-14 (hypothesis 2) was supported by a strong positive correlation with the AQ-10 autism spectrum quotient, r = .81. Discriminant validity (hypothesis 3) was also supported by a strong negative correlation with the EQ-Short, r = -.75. However, the measure did not meet inferential criteria for internal consistency (hypothesis 1), and confirmatory factor analysis found a poor fit of the proposed three-factor model (hypothesis 4) to the data. A cut-off score of 14/42 provided adequate sensitivity (95%) to detect participants with self-reported ASD diagnoses, but not adequate specificity (70%). %), suggesting a very high rate of false positives should be expected if relying on RAADS-14 scores alone to interpret presence of ASD. In sum, our results do not provide sufficient evidence of reliability and validity to support the use of the RAADS-14 with the New Zealand population. We provide suggestions for refinement of the RAADS-14 that may lead to increased reliability and validity.

Keywords: Autism spectrum disorder, ASD, RAADS-14, screening, validity

Autism in Aotearoa: Is the RAADS-14 a valid tool for a New Zealand population

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition impacting the behaviour, communication and perception of the world of affected individuals. Understanding ASD as a spectrum, as reflected in the latest Diagnostic and Statistical Manual of the American Psychiatric Association, the DSM-5 (American Psychiatric Association, 2013), is important as each person is affected differently, with varying degrees of severity. ASD can affect a person of any ethnicity, culture or sociodemographic status, with wide-ranging implications for themselves, their families and their communities (Samadi & McConkey, 2011). In New Zealand, prevalence is believed to mirror the US rate of 1/59 individuals (Baines & Yates, 2018; Baio, 2018). Although, as a neurodevelopmental disorder, ASD is inevitably present from early life, it is not always detected in childhood. Screening adults for unrecognised ASD is important - other comorbid conditions may develop or be presumed, confusing presentation or diagnosis and delaying access to support (Aggarwal & Angus, 2015; Bargiela, Mandy, & Steward, 2016). To date, the cause of ASD remains unknown, and no pathognomonic markers, nor a "cure" have yet been identified (Happé, Ronald, & Plomin, 2006). However, adverse impacts can be improved with early, targeted intervention, necessitating reliable, valid assessment and diagnostic measures (Landa, 2018).

The importance of screening for ASD

In any population, diagnosing ASD is a complex, multi-stage, multidisciplinary process, both time-consuming and costly (Horlin, Falkmer, Parsons, Albrecht, & Falkmer, 2014). Yet, despite evidence of variation in symptom presentation by ethnicity or culture (Soto et al., 2015), there is a stark lack of population-specific ASD knowledge. For example, in New Zealand, ASD prevalence is inferred from a combination of international data and a limited sample informing

the Ministry of Health general health survey (Ministry of Health, 2017). In 2011, McClintock and Fraser reported that no screening, assessment or diagnostic instruments had been validated for the New Zealand population. Literature searches suggest this has not changed in the intervening years. Official assessment guidelines are largely aspirational as, in reality, the availability and structure of expert services in any one region varies widely (Thabrew & Eggleston, 2018). Long wait times exacerbate an already stressful and, for many, unsatisfactory process (Eggleston, Thabrew, Frampton, Eggleston, & Hennig, 2019).

Screening measures for ASD, among other conditions, are a convenient, low-cost and usually brief way for individuals or practitioners to identify or rule out a need for further investigation and are vital where resources are limited (Durkin et al., 2015). Many such instruments exist and can expedite identification and provision of essential support for individuals whose difficulties might otherwise go undetected. Although screening is encouraged in many medical and psychological domains, however, it is important that the measures themselves are reliable and valid, and that they are properly administered and interpreted (Marlow, Servili, & Tomlinson, 2019). This is particularly important in respect of ASD in light of evidence of cross-cultural differences in manifestation and recognition of symptoms (Elsabbagh et al., 2012; Mandell et al., 2009), demanding careful evaluation of the performance of instruments developed for one population when used with another (Donohue, Childs, Richards, & Robins, 2019; Elsabbagh et al., 2012). Screening measures are not used for diagnosis alone but are also convenient research tools. Brevity, accessibility and ease of administration make these a popular choice given what are often tight financial constraints, and to minimise participant fatigue. Research findings, which often lead to changes in policy and practice, must rest on valid foundations. However, despite this clear imperative, Harrison, Slane, Hoang, &

Campbell (2017) report that, to date, only a minority of studies in the ASD field have used measures with strong, established psychometric properties. This underscores a need to identify workable, but valid measures for any population of interest, whether for clinical or academic purposes. In particular, a measure must be sufficiently sensitive to detect true positive results – for example, the RAADS-14 must be able to accurately detect individuals with ASD. It must also be sufficiently specific to correctly distinguish individuals who do not have ASD, measured by scores falling below the determined cut-off point (Glaros & Kline, 1988).

Evaluating the RAADS-14 screen: Study rationale

Using a New Zealand sample, this study considered reliability and validity of the RAADS-14 (Eriksson, Andersen, & Bejerot, 2013). This is a brief (14-item) self-report screening measure for adults based on the Ritvo Autism and Asperger Diagnostic Scale-Revised (RAADS-R; Ritvo et al., 2011). In developing the RAADS-14, Eriksson et al. (2013) recruited 135 Swedish adults with ASD, a further 508 with varying psychiatric diagnoses - including attention deficit hyperactivity disorder (ADHD) - and 590 non-clinical controls. Scores for each participant were obtained across three domains, each theoretically considered a core component of ASD symptomology – mentalising deficits, social anxiety and sensory sensitivity. The authors reported both excellent internal consistency reliability for the full scale (n = 1,233, α = .9) and power to discriminate between ASD and non-psychiatric groups (area under curve (AUC) .99). Discriminatory power (ASD group vs “other psychiatric”) was found to be moderate to good (AUC .91; .88). A cut-off total score of 14/42, determined by selecting the lowest score corresponding to a true positive rate of 93% or greater during development, yielded .97 sensitivity, and specificity ranging from .46 (against the ADHD sample); to .64 (other psychiatric); and .95 (non-psychiatric). Overall, total scores were significantly higher for the

ASD group than for controls ($p < .001$), with large effect sizes. At the item level, overall scores on each item were lower in control samples ($p < .001$).

In a recent study evaluating ASD brief screening tools, despite remarking on the poorer ability of the RAADS-14 to distinguish other psychiatric conditions, especially ADHD, Wigham et al. (2019) acknowledged its high levels of sensitivity and specificity, compared to a non-psychiatric population. The RAADS-14 was selected for present purposes on the strength of the reported psychometric properties and its recency, meaning its development benefited from comparatively current research on ASD item content.

Aims

This study evaluated the reliability and validity of the RAADS-14, to test and extend reporting of its psychometric properties. Scores from a New Zealand sample were obtained on the RAADS-14 and on a conceptually similar measure, the AQ-10, a short version of the Autism Spectrum Quotient (Allison, Auyeung, & Baron-Cohen, 2012). Participants also completed a brief empathy measure, the EQ-Short (Wakabayashi et al., 2006), drawing on theory that individuals with ASD are comparatively low in empathy, compared to the general population (Baron-Cohen & Wheelwright, 2004).

In particular, it was hypothesised that:

1. The internal consistency reliability of the RAADS-14 overall, and that of each of the three subscales (mentalising deficits, sensory reactivity, and social anxiety) would be high ($\omega > .8$);
2. Total scores on the RAADS-14 would have a positive and substantial correlation with scores on the AQ-10 ($r > .3$);

3. Total scores on the RAADS-14 would negatively correlate with scores on the EQ-Short ($r < -.2$)[1];
4. A single-order three-factor model (as reported in the original validation study, see also Figure 1) would explain the covariance between RAADS-14 items;
5. The RAADS-14 total cut-off score (14/42) would permit identification of ASD in individuals with confirmed ASD diagnoses (expected minimum sensitivity .9; minimum specificity .85), closely approximating the findings by Eriksson et al. (2013).

Preregistration

Motivated by the potential research, policy and practice improvement implications of these findings to further understanding ASD measurement in New Zealand, this study was preregistered with the Open Science Framework (OSF), available at https://osf.io/e7hbd. Transparency through preregistration, including data-sharing for replication and accountability, is an invaluable method to counter the said "replication crisis" in psychology (Flake & Fried, 2019) and, as Hussey and Hughes (2018) explain, is particularly important in the investigation of measurement tools. The study was also submitted as a registered report to the European Journal of Psychological Assessment, which accepted publication in principle on the basis of the Stage 1 manuscript on 12 September 2019 after peer-review. The final analysis code and anonymised data are now also accessible on OSF (https://osf.io/4szdg/).

[1] Due to typographical error, this was incorrectly stated as "r > -.2" in the Stage 1 manuscript.

Method

Participants

Participants were recruited using non-probability sampling methods. Two subsamples were formed from the total participant group, based on self-disclosed ASD diagnostic status – a "general population group" and an "ASD group". The aim was to collect a sample of data from 385 individuals, to ensure a 95% confidence interval no wider than .2 when estimating correlations (for hypotheses 2 and 3). However, a recruitment target of at least 420 participants was specified to allow for the possibility that some participant data would need to be excluded, as per the exclusion criteria outlined below. To permit testing of hypothesis 5, an absolute minimum of 43 participants was required within each of the two subsamples (ASD and general population), permitting a margin of error for sensitivity and specificity of less than 15%.

Data collection was to continue until at least 420 responses in Qualtrics with a status of "Finished" were collected. Of this total pool, a minimum of 50 respondents (allowing for the possibility that some participants might need to be excluded) was specified as required to have declared an existing ASD diagnosis, and a minimum 50 who declared no diagnosis, before collection would stop. Satisfaction of these three criteria was necessary to ensure sufficient numbers existed overall, as well as in the relevant subgroups, to permit the planned analyses. Once data collection began, participant numbers were checked weekly (between 9-11am (NZT) on Fridays) and ceased when each of these specified conditions were met.

To partake, an individual was required to live in New Zealand, or be a citizen, aged at least 18, and capable of giving informed consent. No children, or others for whom caregivers would need to consent or complete the survey on their behalf, were eligible.

Advertising was conducted primarily through social media platforms such as New Zealand university and psychology-related Facebook groups and other forms of electronic communication and supplemented by targeted advertisements on Twitter and Facebook. It also included newsletter mailouts by domestic charitable organisations Autism New Zealand and Altogether Autism. No personal reward was offered but participants were informed that, for every completed survey, the University would donate NZ$1 to each of these two organisations.

Data collection began on 16 September 2019 and ceased on 1 November 2019 as the stopping criteria were met. Incomplete surveys started within this timeframe were permitted to be finalised, the last of these being on 7 November. The total number of completed surveys was 440, of which 50 participants had self-reported a diagnosis of ASD. The final sample size, after applying specific exclusion criteria (outlined below), was 387 (female = 302; modal age bracket 36-65 years), of whom 41 had declared an ASD diagnosis. For a detailed description of participants' demographic characteristics, see the Supplementary Materials.

Procedure

All prospective participants were provided with an electronic link to a Qualtrics survey (Qualtrics, 2019), accessible on desktop or mobile platforms, containing a detailed information sheet and consent form. Confirmation of consent launched the survey, which participants were advised would take approximately 10 minutes to complete.

Minimal demographic data was collected in the preliminary section of the survey. Participants indicated their age group, self-identified gender and whether they were either a New Zealand citizen, or otherwise resident in New Zealand. To permit assignment to either the general population or ASD group, necessary to test hypothesis 5, participants were asked

whether they had ever had a formal diagnosis of ASD or related condition, including ASD, autism, Asperger's syndrome or pervasive developmental disorder - not otherwise specified.

Data exclusions

Anyone under 18, or not either a New Zealand citizen or living in New Zealand, was directed out of the survey by their responses to the demographic questions and was unable to progress further. Other responses were excluded in their entirety where one or more of the following occurred:

- total completion time (as measured in Qualtrics) was below 2 minutes, indicating responses were selected randomly;
- the response to an attention check question ("*At points in my life, I have drunk water*"), embedded in the EQ-Short items, was anything other than "strongly agree", suggesting the participant was not attending to item content;
- answers suggestive of response set, determined by selection of the same response alternative for every statement within any one of the three main study scales – the AQ-10, EQ-Short, or RAADS-14; and
- responses assigned a status by Qualtrics of 1 (preview), 2 (test), 8, 9, or 12 (possible spam or duplicate responses).

A missed answer or answers to a small number of questions did not automatically exclude an entire survey, however a minimum of 75% of items (35 of the total 46, excluding the attention check – of which 14 from the RAADS-14, 10 from the AQ-10, and 22 from the EQ-Short) were required to be answered for a participant to be included in the sample.

Ethics

This study was approved by the [REDACTED FOR ANONYMITY] University Human Ethics Committee.

Measures

The substantive survey questions were identical for every participant, set out in three separate blocks of questions for each of the RAADS-14, AQ-10 and EQ-Short (see Appendix). Question order and response alternatives matched those in the original measures.

RAADS-14

The RAADS-14 is a 14-item self-report screening inventory based on the RAADS-R. Each question is mapped to one of three domains, revealed during the development phase to account for most of the total variance in scores and comprising the subscales targeted in the present study – mentalising deficits (items 1, 4, 9, 11, 12, 13, 14[2]); social anxiety (3, 5, 6, 8) and sensory reactivity (2, 7, 10). Each item also relates to a particular DSM-5 criterion for ASD (e.g. mentalising deficits: item 4 – *"it is difficult to figure out what other people expect of me"* relates to DSM-5 criterion A2; social anxiety: item 5 – *"I often don't know how to act in social situations"* relates to DSM-5 criterion A1; sensory reactivity: item 7 - *"when I feel overwhelmed by my senses, I have to isolate myself to shut them down"* relates to DSM-5 criterion B4).

Responses are provided on a 4-point rating scale – True now and when I was young (3); True only now (2); True only when I was younger than 16 (1); and Never true (0), except item 6 which is reverse worded, and scored accordingly. Although the original study reported a 3-factor structure, this is on the basis that each of these factors contributes to a single overarching

[2] In the stage 1 manuscript, item 14 was omitted from this list in error. This typographical error was identified and corrected prior to any analysis of data.

construct of ASD. As such, participants' total scores (possible range – 0-42) were used in respect of each of the hypotheses, and individual scale scores calculated to test hypothesis 3 (see Table 3, Eriksson et al., 2013).

AQ-10

Like the RAADS-14, the AQ-10 is an abbreviated, self-report version of an existing test - the Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). Of the ten items of the AQ-10, two are taken from each of the five sub-scales of the original AQ (attention to detail; attention switching; communication; imagination; and social). Participants choose one of 4 response options per item which, like the RAADS-14, are "definitely agree", "slightly agree", "slightly disagree" and "definitely disagree". Scoring differs however - each attracts either 0 or 1 point. Either of the "agree" options on items 1, 7, 8 and 10, or either of the "disagree" responses on items 2, 3, 4, 5, 6 and 9 attracts 1 point. A score of 6 or more indicates specialist referral should be considered (Allison et al., 2012).

EQ-Short

Drawing on the empathising-systemising theory of autism (Baron-Cohen, 2009), the purpose of administering the EQ-Short is to enhance the understanding of the RAADS-14 performance, specifically to consider the extent of its discriminant validity. Response options are the same as for the AQ-10, and each of the 22 items is scored from 0 to 2. For positively worded items (1, 2, 6, 8, 9, 10, 12, 13, 14, 15, 16, 18, 19, 10, 21, 22)[3], 0 points are awarded for a "disagree" response, 1 is awarded for "slightly agree" and 2 for "strongly agree". Negatively

[3] Items 1, 3, 11, 13, 14, 15, 21, 22, 26, 28, 29, 34, 35, 36, 38, 39 from the original EQ scale.

worded items (3, 4, 5, 7, 11, 17)[4] score 0 for an "agree" response, 1 for "slightly disagree" and 2 for "strongly disagree".[5] Total scores then range from a possible 0 to 44.

Data Analysis

All analyses were performed in R (R Core Team, 2013), after data was "cleaned" to identify and evaluate missing data points and exclude incomplete or invalid responses, as per the criteria outlined above. Where participants missed fewer than 11 items to any of the three main study scales (RAADS-14, AQ-10 and EQ-Short), responses to those items were imputed using single expectation-maximisation imputation, implemented in the Amelia package in R (Honaker, King, & Blackwell, 2011). Hypotheses 1-4 were tested using the full sample, whereas hypothesis 5 required comparison of scores between the general population and ASD groups.

Hypothesis 1: Reliability of the RAADS-14

To assess internal consistency reliability of the RAADS-14, overall, and by subscale, McDonald's omega (ω) was computed using the MBESS package in R (Kelley & Lai, 2012), in accordance with the preregistration. Hypothesis 1 was to be considered supported if the observed ω for the total scale, and for each for the three subscales, was greater than .8, using Nunnally's rule of thumb for research purposes (Nunnally, 1978). Reported together with a confidence interval, ω indicates variability in the estimation, and the degree of confidence in the reliability of the RAADS-14 (Dunn, Baguley, & Brunsden, 2014).

---

[4] Items 4, 8, 9, 12, 18, 31 from the original EQ scale.

[5] In the present study, an attention check question was embedded within the EQ-Short questions, between items 13 and 14, changing the numbering from 15. This is important to the analysis, and should be noted for replication purposes, as it is item 18, rather than 17, that must be reverse coded.

Hypotheses 2 and 3: convergent and discriminant validity of RAADS-14

A Pearson product-moment correlation coefficient was calculated to assess the magnitude and direction of the relationship between the total RAADS-14 scores and the total scores on the AQ-10 (hypothesis 2, convergent validity) and EQ-Short (hypothesis 3, discriminant validity), respectively.

Hypothesis 2 was to be considered supported if the correlation between scores on the RAADS-14 and AQ-10 was positive, statistically significant ($p < .05$, 2-tailed), and greater than .3. Hypothesis 3 was considered supported providing a negative, statistically significant ($p < .05$, 2-tailed) correlation was observed between RAADS-14 and EQ-Short scores ($r < -.2$)[6].

Hypothesis 4: RAADS-14 three-factor structure

Hypothesis 4 was tested by specifying a single-order three-factor model. The items loading on each domain specified by the model are set out in Figure 1.

---

[6] Although this required magnitude was specified in the list of hypotheses, it was omitted from the method section in the Stage 1, preregistered report. It has been added here for clarity and completeness.

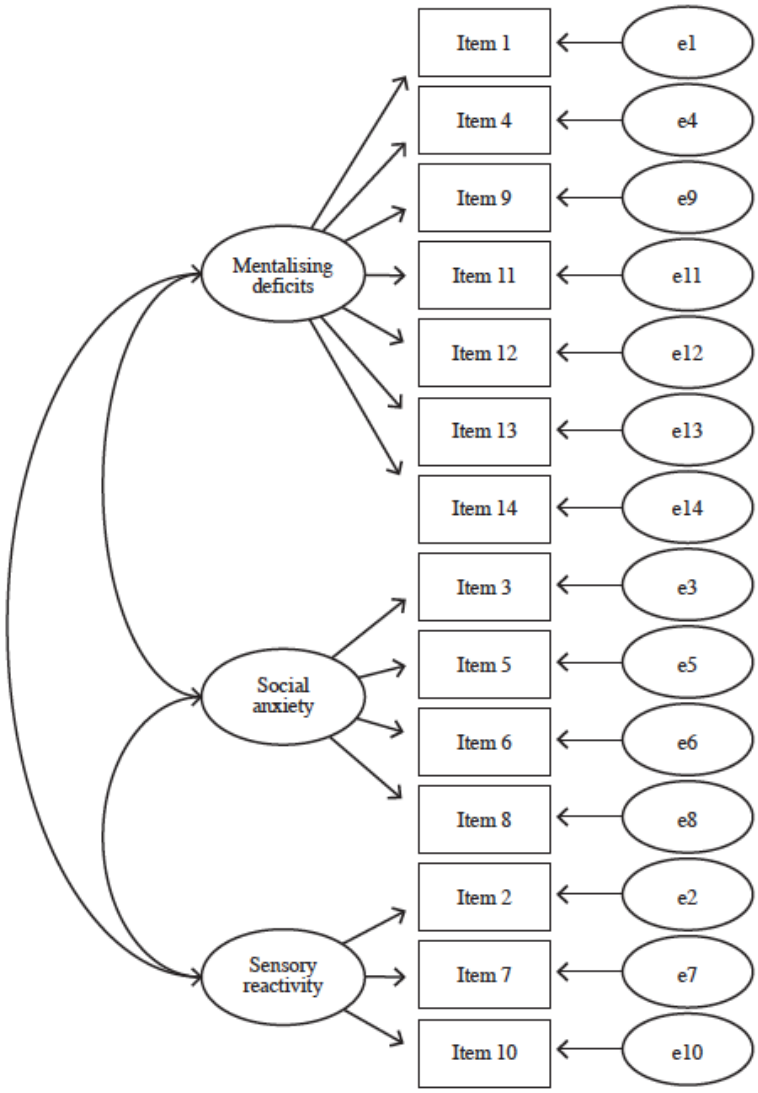

Figure 1: SEM path diagram – RAADS-14 single order model

Parameter estimates were to be considered statistically significant if the obtained *p* values were less than .05 (2-tailed). Diagonally weighted least squares were used to estimate the model, together with robust standard errors and a Satorra-Bentler-scaled test statistic.

The fit of the model proposed, that is its ability to reproduce the data and its consistency, was assessed by calculation of the following statistics, interpreted according to guidance outlined by Hu and Bentler (1999). Fit would be considered supported only where both the root mean

square error of approximation (RMSEA) was less than .06, and the comparative fit index (CFI) was greater than .95 (Hu & Bentler, 1999).

Hypothesis 5: Sensitivity and specificity of the RAADS-14

Sensitivity of the measure was estimated by calculating the percentage of individuals with an existing ASD diagnosis who received total scores equal to or greater than 14/42. Specificity was estimated from the percentage of individuals without an existing diagnosis who received scores lower than 14 on the RAADS-14. What constitutes "good" or "high" sensitivity or specificity is context dependent, both in terms of a test itself and the implications of reliance on either a false positive or false negative result (Glaros & Kline, 1988; Lalkhen & McCluskey, 2008). Hypothesis 5 was to be considered supported if the estimated sensitivity was equal to or greater than .90, and the estimated specificity equal to or greater than .85.

Relevance of the RAADS-14 for Aotearoa/New Zealand

The preregistered criteria for this study specified that if any of hypotheses 1, 2 or 3 were not supported by the findings, the RAADS-14 would be considered to have demonstrated insufficient reliability and/or validity for use with a population in Aotearoa/New Zealand. If hypotheses 1-3 were supported, but not either hypothesis 4 and/or 5, the results would be considered equivocal. If all five hypotheses were supported, the RAADS-14 was to be considered a good fit for further research with a New Zealand population, although specific attention to its use with Māori would be required.

Results

Participant scores on each measure, overall and by subsample group, are described at Table 1.

Table 1.

Descriptive Statistics: RAADS-14, AQ-10, EQ-Short

	Full sample		ASD group		General population group	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
RAADS-14	11.99	11.26	29.86	9.29	9.85	9.43
AQ-10	2.68	3.41	6.99	2.64	2.98	2.35
EQ-Short	26.17	11.57	12.75	9.18	27.78	10.76

Note. Scoring: RAADS-14: 0-42; AQ-10: 0-10; EQ-Short: 0-44. *M* and *SD* represent mean and standard deviation, respectively

Reliability of the RAADS-14

Our Stage 1 manuscript stated that we would estimate reliability using McDonald's omega. We did not explicate which form of omega we would report, but did include a citation to Dunn et al. (2014), which advocates the omega of McDonald (1999)—i.e., hierarchical omega. As such, we report hierarchical omega estimates here. Total scale reliability for the RAADS-14 was estimated at ω = .89, 95% CI [.87, .91]), exceeding the preregistered inferential criteria (ω > .80). The results also suggest reliability of the mentalising deficit subscale (ω = .84, 95% CI [.79, .87]). However, reliability of the remaining two subscales fell below the preregistered threshold (social anxiety ω = .79; 95% CI [.76, .82]; sensory reactivity (ω = .65, 95% CI [.58, .72]). Accordingly, the first research hypothesis is not supported on the basis of the strict inferential criteria. Further information on the computation of reliability estimates, and analyses for the AQ-10 and EQ-Short, are provided in the Supplementary Materials.

Convergent and discriminant validity of RAADS-14

A strong positive correlation was observed between participants' total scores on the RAADS-14 and the AQ-10 ($r(385) = .81$, $p < .001$, 95% CI [.77, .84]). Total scores on the

RAADS-14 were strongly negatively correlated with those on the EQ-Short ($r(385) = -.75$, $p < .01$, 95% CI [- .79, - .71]. Hypotheses 2 and 3 are therefore supported by these results.

RAADS-14 three-factor structure

Confirmatory factor analysis (CFA) was conducted to test the fit of the data to the three factor, single order model. Three latent variables – mentalising deficits, social anxiety and sensory reactivity – are proposed in this model (see Figure 1). Here, the fit statistics showed poor fit of the data to the model: (χ^2 (74) = 243.05, $p < .001$; comparative fit index (CFI) = .92; root mean square error of approximation (RMSEA) = .08, 90% CI [.07, .09]). Neither the RMSEA nor the CFI met the specified criteria (RMSEA < .06, and CFI > .95). Accordingly, hypothesis 4 is not supported. Figure 2 depicts the fully standardised latent variable loadings and model covariances obtained.

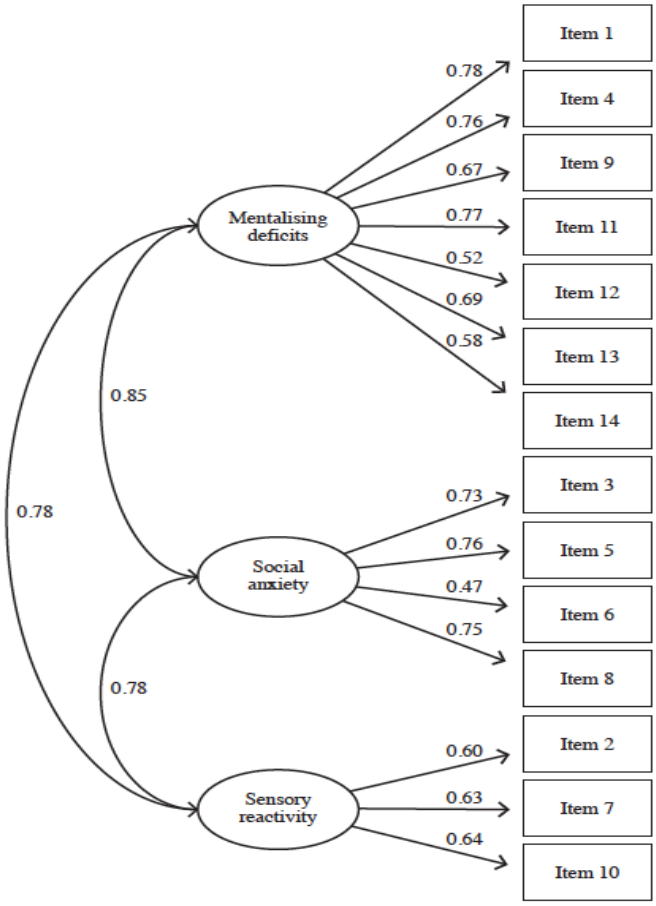

Figure 2. RAADS-14 SEM path diagram with fully standardised latent variable loadings and covariances.

Sensitivity and specificity of the RAADS-14

In this sample, of the 41 individuals with a declared ASD diagnosis, 39 scored at or higher than the cut-off score of 14 (95.12%). Sensitivity of the RAADS-14 is therefore estimated at .95 (CI [.93, .97]). However, only 70.26% of participants in the general population group scored below 14, 102 of these individuals reaching or exceeding the cut-off score. Estimated specificity of the RAADS-14 is therefore .7 (95% CI [.66, .75]). As such, despite a finding of sensitivity above the preregistered $\geqslant$ .90, specificity fell short of the $\geqslant$ .85 required to support hypothesis 5 in this study.

Taking into account the reported ASD population prevalence estimate for New Zealand of 1/59 (Baines & Yates, 2018), this results in an estimated positive predictive value (PPV) of the RAADS-14 of 5.26% and a negative predictive value (NPV) of 99.88%.

Discussion

Overall, in accordance with the explicit preregistered inferential criteria, the findings in this study did not demonstrate sufficient reliability or validity of the RAADS-14 to recommend its use with the New Zealand population. Although the findings suggested convergent and discriminant validity of the measure, by virtue of a strong positive correlation with observed with the AQ-10, and a strong negative correlation with the EQ-Short (hypotheses 2 and 3), the remaining three hypotheses were not supported.

Internal consistency reliability of the total scale and mentalising deficits subscale did exceed the specified threshold ($\omega = .8$), but reliability of the two remaining subscales – social anxiety and sensory reactivity – fell short (meaning that hypothesis 1 was not supported). The CFA results identified difficulties with the sensory reactivity subscale in particular and, overall, the fit statistics – as a gauge of its consistency and ability to reproduce the data - suggested that the 3-factor model based on the original study findings was a poor fit, and unlikely to adequately explain relationships between RAADS-14 items (hypothesis 4).

In respect of hypothesis 5, whilst the results suggest the RAADS-14 is sensitive – using the cut-off score for detection of ASD of 14 specified by the test developers, it accurately identified 95% of participants in the sample who self-reported a prior ASD diagnosis - nearly 30% of participants in the general population group also scored at or above this cut-score (70.26% specificity). As such, were the RAADS-14 to be used for identifying ASD in the absence of other corroborating information, a very high rate of false positives should be

expected. Caution must therefore be exercised before relying on RAADS-14 scores to interpret presence of ASD alone. This said, diagnosis of any psychological disorder should not take place in isolation from other contextual information and collateral sources of data, and where the purpose of any screening measure is to isolate individuals at higher risk for onward referral and assessment, testing with the RAADS-14 will likely achieve this aim, albeit with a high number of false positives.

These results are not surprising, given the results of previous research highlight the need for careful attention to the performance of any instrument developed and standardised for one population when it is used for another (Donohue, Childs, Richards, & Robins, 2019; Elsabbagh et al., 2012). This is particularly true in respect of instruments to screen for or diagnose ASD given reported cultural differences in the manifestation and recognition of its symptoms (Elsabbagh et al., 2012; Mandell et al., 2009). The findings of this study support a need for local validation of measures used to screen for and diagnose ASD in Aotearoa/New Zealand.

Limitations and directions for future research

Despite efforts to purposively recruit individuals eligible for inclusion in the ASD group for this study and strict adherence to the recruitment stopping rule, after applying the exclusion criteria, the eventual number in this group was 41, two below the desired minimum of 43 specified at preregistration. Of the 56 individuals who declared a diagnosis of ASD when the survey closed, 2 were excluded having given the same response to all items on the RAADS-14. A further 13 were excluded by virtue of the required "strongly agree" response to the attention check question (*"At points in my life, I have drunk water"*). It is possible that some participants who were attending to the questionnaire considered "agree" to be a sufficient response to the question. The impact of applying a more lenient exclusion rule here is detailed in the

Supplementary Materials file (but does not alter the overall conclusions). Further, women greatly outnumbered men or gender diverse individuals, in both study subsamples, comprising 65.9% of the ASD group and 80.2% of the general population group. This can be expected to impact generalisability of results, particularly given higher reported base rates of ASD diagnosis in males compared with females globally.

Measurement difficulties arising from allocation to the ASD or general population group dependent on self-declared status alone are acknowledged. In the former group, diagnostic status was not independently verified and, in the latter, individuals who would not meet ASD diagnostic criteria could not be distinguished from others who might, but had never been diagnosed. This introduces additional uncertainty to the conclusions relating to the sensitivity and specificity of the RAADS-14. To increase accuracy and confidence in the findings, particularly with respect to sensitivity and specificity, future studies could seek to obtain samples more representative of the demographic characteristics of the broader population. It may also be beneficial to include a consistent diagnostic measure of participants such as a diagnostic interview, as an additional step to enhance confidence that criteria for ASD are in fact met.

It is acknowledged, with hindsight, that the inferential criteria specified for concluding sufficient reliability or validity of the RAADS-14 for use in Aotearoa/New Zealand may have been excessively high, in particular as relates to the internal consistency reliability of the measure. On these results, the full scale was found to be reliable according to the inferential criterion specified, as was reliability of the mentalising deficits subscale. However, the estimates of reliability for the sensory reactivity and social anxiety subscale reliability were lower, meaning that the overall inferential criteria for concluding adequate reliability were not met. However, the degree to which the reliability of the subscales matters depends on context - for

researchers or clinicians only making use of the total scale (as is likely to be the case when screening for ASD), the reliability of individual subscales is largely irrelevant. For this reason, it could be argued that we should not have specified inferential criteria that depended on the reliability of individual subscales.

To date, no instrument for screening (or diagnosing) ASD has been validated for the population in Aotearoa/New Zealand. In these circumstances, and given the findings of reliability of the full scale with the current sample, further research on how the RAADS-14 may be modified to enhance its psychometric properties for the New Zealand population may be justifiable. Our study provides two important directions in that respect. First, item 6 ("*I can chat and make small talk with people"*) had the lowest standardized factor loading of all the items. This is the only negatively worded item in the scale, and its difference in wording and lower factor loading may mean that it makes a less substantial contribution to the reliability of the scale than other items. Replacing it with an appropriate positively worded item might result in a small improvement to reliability. Second, while the social anxiety and sensory reactivity subscales demonstrated relatively weak reliability, this is likely due simply to their small number of items (four and three items, respectively). This could be addressed in future research by increasing the number of items in these subscales, or simply by using only the total score (and not subscale scores) for substantive purposes.

Conclusions

In this study, the results do not support the reliability, construct validity or specificity of the RAADS-14. Accordingly, the RAADS-14 cannot be said overall to be sufficiently reliable or valid for use as a measure to screen for ASD in Aotearoa/New Zealand without further

evaluation or modification. However, there is as yet no other measure than has been established as reliable or valid in this country. It is therefore important to build on these findings to improve understanding of ASD in Aotearoa/New Zealand, as well as differences in detection and diagnosis globally. Further research is needed and may take several forms. Firstly, studies could extend these findings, seeking to remedy limitations identified here. Secondly, consideration should be given to whether the RAADS-14 could be modified to improve its psychometric properties in this population. Third, validation research could be conducted with other measures of ASD to examine their reliability and validity amongst people in Aotearoa/New Zealand.

Conflict of interest declaration

The authors declare no conflict of interest.

References

Aggarwal, S., & Angus, B. (2015). Misdiagnosis versus missed diagnosis: Diagnosing autism spectrum disorder in adolescents. *Australasian Psychiatry*, *23*(2), 120–123. https://doi.org/10.1177/1039856214568214

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "red flags" for autism screening: The short Autism Spectrum Quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(2), 202-212.e7. https://doi.org/10.1016/j.jaac.2011.11.003

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed*. https://doi.org/10.1176/appi.books.9780890425596

Baines, P., & Yates, A. (2018). Autism spectrum disorder in Aotearoa New Zealand: Strategies for teachers. *Teachers' Work*, *15*(2), 89–93. https://doi.org/10.24135/teacherswork.v15i2.261

Baio, J. (2018). Prevalence of autism spectrum disorder among children aged 8 years. *MMWR. Surveillance Summaries*, *67*. https://doi.org/10.15585/mmwr.ss6706a1

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

Baron-Cohen, S. (2009). Autism: The empathizing–systemizing (E-S) theory. *Annals of the New York Academy of Sciences*, *1156*(1), 68–80. https://doi.org/10.1111/j.1749-6632.2009.04467.x

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.

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175.

Donohue, M. R., Childs, A. W., Richards, M., & Robins, D. L. (2019). Race influences parent report of concerns about symptoms of autism spectrum disorder. *Autism*, *23*(1), 100–111. https://doi.org/10.1177/1362361317722030

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Durkin, M. S., Elsabbagh, M., Barbaro, J., Gladstone, M., Happé, F., Hoekstra, R. A., … Shih, A. (2015). Autism screening and diagnosis in low resource settings: Challenges and opportunities to enhance research and services worldwide. *Autism Research*, *8*(5), 473–476. https://doi.org/10.1002/aur.1575

Eggleston, M. J. F., Thabrew, H., Frampton, C. M. A., Eggleston, K. H. F., & Hennig, S. C. (2019). Obtaining an autism spectrum disorder diagnosis and supports: New Zealand parents' experiences. *Research in Autism Spectrum Disorders*, *62*, 18–25. https://doi.org/10.1016/j.rasd.2019.02.004

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcín, C., … Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research*, *5*(3), 160–179. https://doi.org/10.1002/aur.239

Eriksson, J. M., Andersen, L. M., & Bejerot, S. (2013). RAADS-14 Screen: Validity of a screening tool for autism spectrum disorder in an adult psychiatric population. *Molecular Autism*, *4*(1), 49. https://doi.org/10.1186/2040-2392-4-49

Flake, J. K., & Fried, E. I. (2019). *Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them* [Preprint]. https://doi.org/10.31234/osf.io/hs7wm

Glaros, A. G., & Kline, R. B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of Clinical Psychology*, *44*(6), 1013–1023. https://doi.org/10.1002/1097-4679

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience*, *9*(10), 1218–1220. https://doi.org/10.1038/nn1770

Harrison, A. J., Slane, M. M., Hoang, L., & Campbell, J. M. (2017). An international review of autism knowledge assessment measures. *Autism*, *21*(3), 262–275. https://doi.org/10.1177/1362361316638786

Horlin, C., Falkmer, M., Parsons, R., Albrecht, M. A., & Falkmer, T. (2014). The cost of autism spectrum disorders. *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0106552

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hussey, I., & Hughes, S. (2018). *Hidden invalidity among fifteen commonly used measures in social and personality psychology*. https://doi.org/10.31234/osf.io/7rbfp

Kelley, K., & Lai, K. (2012). MBESS. R package (Version 3.3.2). Retrieved from http://CRAN.R-project.org/package=MBESS

Lalkhen, A. G., & McCluskey, A. (2008). Clinical tests: Sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain*, *8*(6), 221–223. https://doi.org/10.1093/bjaceaccp/mkn041

Landa, R. J. (2018). Efficacy of early interventions for infants and young children with, and at risk for, autism spectrum disorders. *International Review of Psychiatry*, *30*(1), 25–39. https://doi.org/10.1080/09540261.2018.1432574

Mandell, D. S., Wiggins, L. D., Carpenter, L. A., Daniels, J., DiGuiseppi, C., Durkin, M. S., … Kirby, R. S. (2009). Racial/ethnic disparities in the identification of children with autism spectrum disorders. *American Journal of Public Health*, *99*(3), 493–498. https://doi.org/10.2105/AJPH.2007.131243

Marlow, M., Servili, C., & Tomlinson, M. (2019). A review of screening tools for the identification of autism spectrum disorders and developmental delay in infants and young children: Recommendations for use in low- and middle-income countries. *Autism Research: Official Journal of the International Society for Autism Research*, *12*(2), 176–199. https://doi.org/10.1002/aur.2033

McClintock, J. M., & Fraser, J. (2011). *Diagnostic instruments for autism spectrum disorder A brief review.* Wellington: New Zealand Guidelines Group, for the Ministries of Health and Education. Retrieved from https://www.health.govt.nz/system/files/documents/publications/asd_instruments_report.pdf .

Ministry of Health. (2017). *New Zealand Health Survey: 2016/2017*. Retrieved from https://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/surveys/new-zealand-health-survey

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*(3), 137-141. http://dx.doi.org/10.1027/1864-9335/a000192

Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill.

Perry, J. L., Nicholls, A. R., Clough, P. J., & Crust, L. (2015). Assessing model fit: Caveats and recommendations for confirmatory factor analysis and exploratory structural equation modeling. *Measurement in Physical Education and Exercise Science*, *19*(1), 12–21. https://doi.org/10.1080/1091367X.2014.952370

Qualtrics. (2019). Qualtrics XM (Version XM). Retrieved from https://www.qualtrics.com/au

R Core Team. (2013). *R: The R project for statistical computing*. Retrieved from https://www.r-project.org/

Revelle, W. (2018). psych: Procedures for personality and psychological research (Version 1.8.12) [R, R]. Retrieved from https://CRAN.R-project.org/package=psych

Revelle, W., & Condon, D. (2018). *Reliability from alpha to omega: A tutorial*. https://doi.org/10.31234/osf.io/2y3w9

Ritvo, R. A., Ritvo, E. R., Guthrie, D., Ritvo, M. J., Hufnagel, D. H., McMahon, W., … Eloff, J. (2011). The Ritvo autism Asperger diagnostic scale-revised (RAADS-R): A scale to assist the diagnosis of autism spectrum disorder in adults: An international validation study. *Journal of Autism and Developmental Disorders*, *41*(8), 1076–1089. https://doi.org/10.1007/s10803-010-1133-5

Samadi, S. A., & McConkey, R. (2011). Autism in developing countries: Lessons from Iran. *Autism Research and Treatment*, *2011*. https://doi.org/10.1155/2011/145359

Soto, S., Linas, K., Jacobstein, D., Biel, M., Migdal, T., & Anthony, B. J. (2015). A review of cultural adaptations of screening tools for autism spectrum disorders. *Autism*, *19*(6), 646–661. https://doi.org/10.1177/1362361314541012

Thabrew, H., & Eggleston, M. (2018). Spectrum of care: Current management of childhood autism spectrum disorder (ASD) in New Zealand. *Australasian Psychiatry*, *26*(3), 294–298. https://doi.org/10.1177/1039856217716290

Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J., Fine, D., … Weil, L. (2006). Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short). *Personality and Individual Differences*, *41*(5), 929–940. https://doi.org/10.1016/j.paid.2006.03.017

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*, *23*(2), 287–305. https://doi.org/10.1177/1362361317748245