PSYCHOGERIATRICS
The Official Journal of the Japanese Psychogeriatric Society
doi:10.1111/psyg.12946 *PSYCHOGERIATRICS* 2023

ORIGINAL ARTICLE

Establishing conversion of the 16-item Informant Questionnaire on Cognitive Decline in the Elderly scores into interval-level data across multiple samples using Rasch methodology

Quoc Cuong TRUONG,[1] **Katya NUMBERS**,[2] **Carol C. CHOO**,[3] **Adam C. BENTVELZEN**,[2] **Vibeke S. CATTS**,[2] **Matti CERVIN**,[4] **Anthony F. JORM**,[5] **Nicole A. KOCHAN**,[2] **Henry BRODATY**,[2] **Perminder S. SACHDEV**[2] and **Oleg N. MEDVEDEV**[1]

[1]School of Psychology, University of Waikato, Hamilton, New Zealand, [2]Centre for Healthy Brain Ageing (CHeBA), University of New South Wales, Sydney, New South Wales, [3]College of Healthcare Sciences, James Cook University, Townsville, Queensland, [5]Centre for Mental Health, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia and [4]Department of Clinical Sciences Lund, Child and Adolescent Psychiatry, Faculty of Medicine, Lund University, Lund, Sweden

Correspondence: Quoc Cuong Truong and Dr. Oleg N. Medvedev, School of Psychology, University of Waikato Private Bag 3105, Hamilton 3240, New Zealand. Email: quoccuong513@yahoo.com (Q. C. Truong) and oleg.medvedev@waikato.ac.nz (O. N. Medvedev)

Received 14 December 2022; revision received 20 January 2023; accepted 25 January 2023.

Abstract

Background: The 16-item Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE-16) is a well-validated and widely-used measure of cognitive changes (CCs) among older adults. This study aimed to use Rasch methodology to establish psychometric properties of the IQCODE-16 and validate the existing ordinal-to-interval transformation algorithms across multiple large samples.

Methods: A Partial Credit Rasch model was employed to examine psychometric properties of the IQCODE-16 using data ($n = 918$) from two longitudinal studies of participants aged 57–99 years: the Older Australian Twins Study ($n = 450$) and the Canberra Longitudinal Study ($n = 468$), and reusing the Sydney Memory and Ageing Study (MAS) sample ($n = 400$).

Results: Initial analyses indicated good reliability for the IQCODE-16 (Person Separation Index range: 0.82–0.90). However, local dependency was identified between items, with several items showing misfit to the model. Replicating the existing Rasch solution could not reproduce the best Rasch model fit for all samples. Combining locally dependent items into three testlets resolved all misfit and local dependency issues and resulted in the best Rasch model fit for all samples with evidence of unidimensionality, strong reliability, and invariance across person factors. Accordingly, new ordinal-to-interval transformation algorithms were produced to convert the IQCODE-16 ordinal scores into interval data to improve the accuracy of its scores.

Conclusions: The findings of this study support the reliability and validity of the IQCODE-16 in measuring CCs among older adults. New ordinal-to-interval conversion tables generated using samples from multiple independent datasets are more generalizable and can be used to enhance the precision of the IQCODE-16 without changing its original format. An easy-to-use converter has been made available for clinical and research use.

Key words: *cognitive changes, Informant Questionnaire on Cognitive Decline in the Elderly, measurement, Rasch analysis, reliability.*

INTRODUCTION

Cognitive changes (CCs) reported by older adults (self-reported) or their informants (informant-observed) reflect an individual's personal experience of deterioration in cognitive functions.[1–3] CCs are an essential criterion for the diagnosis of mild cognitive impairment (MCI), which may in turn lead to dementia.[4] Recent studies have shown that both self-reported and informant-observed CCs significantly predicted incident dementia.[5,6] Therefore, CCs may be considered the earliest stage of preclinical dementia.[7,8] Furthermore, CCs can be conveniently captured by self-reported or informant-observed questionnaires that are difficult to detect otherwise in the absence of comprehensive neuropsychological testing.[6]

Researchers have recently examined the advantages of self-reported versus informant-observed CCs.[9] Several studies have shown that informant-observed CCs may be a more reliable approximation for predicting concurrent objective cognitive performance and/or future decline in comparison to self-reported CCs.[5,10–12] Particularly, in the longitudinal Memory and Ageing Study (MAS), cognitive changes were measured by the self-reported Memory Complaint Questionnaire (MAC-Q)[1] and the informant-observed Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) over 10 years.[13] Results from the MAS showed that informant IQCODE scores were more accurate in predicting cognitive performance and incident dementia compared to self-reported MAC-Q scores.[5] Moreover, self-reports are subject to various biases such as the individual's mood (e.g. depression, anxiety), personality traits (e.g. under/over complaining behaviours, neuroticism, or conscientiousness), life events, and medications.[14–16] Even though informant CCs may also be affected by the same factors, although perhaps to a lesser extent,[5,6] informant-observed CC reports are increasingly used in clinical settings as individuals with incipient dementia may lose insight about their cognitive changes.[17]

The 16-item IQCODE-16,[18] a short version of the IQCODE, is widely-used as a CC scale with virtually equivalent psychometric properties to the original 26-item version. The IQCODE-16 has demonstrated excellent internal consistency with Cronbach's alphas ranging from 0.93 to 0.97,[18–21] and its scores have been shown to significantly predict incident dementia.[5,22,23] However, although the IQCODE-16 is a well-validated measure, its scores constitute an ordinal scale. The differences between response options of individual IQCODE-16 items (e.g. 1 and 2 vs 2 and 3) may not as accurately reflect the same amount of clinical change as an interval scale whereby the difference between 1 and 2 is the same as the difference between 2 and 3.[24] All 16 items of the IQCODE-16 are scored using Likert-scale responses ranged ordinally from 1 (much improved) to 5 (much worse), and the total score of the IQCODE-16 is the mean of the items. However, each individual item can reflect a symptom with different severity or difficulty levels, and may contribute differently to the overall latent assessment score (i.e. CC levels).[21,25]

Another issue related to the ordinal nature of the IQCODE-16 scores involves conducting parametric statistics, as they violate the arithmetic assumptions of such tests.[26,27] Rasch analysis[28] can help overcome these limitations. It is a specialized statistical approach that is increasingly used in evaluating the reliability and internal validity of clinical psychometric instruments.[21,25,29–33] Rasch methodology has numerous advantages compared to other widely used statistical methods such as Classical Test Theory[34] and Generalizability Theory,[35] which are unable to account for the different contributions of individual items to the overall latent trait.[36] Specifically, these methods do not distinguish between item difficulty based on individual ability.[28,37] A Rasch model is strictly unidimensional and assumes that a person's response to a particular scale item is determined by that item's difficulty and the capacity of the person on the measured trait.[28,37] Rasch analysis can precisely estimate thresholds between response options of the individual scale items and the contribution of each individual scale item to the overall latent trait.[21,38] Such estimations are possible because most respondents score higher on easy items while only a few score high on more difficult items.[39] Therefore, Rasch analysis can eliminate the biases caused by arbitrarily assigning categorical options to individual items, which are then used to compute total scores, as such scores are not mathematically equal to the total.[40]

The Rasch model also involves invariance testing, meaning it checks for differential item functioning (DIF) as a function of personal factors (e.g. age, sex). DIF is useful to test whether a scale will fit equally for every person regardless of individual factors. If DIF is detected, it indicates that the personal characteristics of a respondent are influencing their responses, meaning that the scale items do not work equally well across different groups.[41] Moreover, the results of Rasch analysis can be presented graphically as an item-person threshold distribution that illustrates how the range of a person's ability is covered by the range of item difficulties, which can be useful for detecting potentially significant ceiling or floor effects.[42] In addition, when the sample data are fitted to the Rasch model, the raw scores of an ordinal instrument can be transformed into interval-level data.[43,44] Such interval-transformed data reflect changes on a latent trait the same way as any other interval-level scale (e.g. Celsius temperature). Increased measurement precision from ordinal-to-interval Rasch transformation has been demonstrated in several studies using different clinical measures.[31,32,43] This is the major benefit of the Rasch model over classical test theory methods, which are unable to produce a genuine interval scale.

This methodology has been recently applied to investigate psychometric properties of the IQCODE-16 in a study where ordinal-to-interval transformation algorithms were proposed that enhanced precision up to that of an interval-level scale.[21] However, this study had limited generalizability as the sample was selectively comprised of participants from a small area of Sydney, Australia. Moreover, while the study had the novel feature of testing the DIF according to the informants' age and sex, it has been more common in ageing research to examine the age and sex of the assessed participants (e.g. Tang and colleagues[45]). These limitations need to be addressed to ensure that transformed scores are generalizable across a wider population of older adults. Therefore, the purpose of the current study was to address these limitations by systematically replicating this earlier study within a larger, more diverse sample to extend the robustness and generalizability of these previous findings.

The first aim of the current study was to replicate Truong and colleagues' previous work applying Rasch methodology to IQCODE-16 scores from informants of participants in the Sydney MAS,[21] as replication studies are important to scientific inquiry and can further enhance the precision of the IQCODE-16. The current study will also allow us to investigate whether the recently developed Rasch model for the IQCODE-16 is generalizable across other samples of participants. Other aims were to investigate invariance (i.e. potential DIF) across the different cohorts and participants and, if necessary, to convert the IQCODE-16 scores into interval-level data. Additionally, we aimed to examine the difference between the new and old conversion algorithms of the IQCODE-16.

METHOD

Participants

Participants' scores for the current study were extracted from two longitudinal studies: the Older Australian Twins Study (OATS),[46] and the Canberra Longitudinal Study (CLS).[47,48] Both studies received ethical approvals permitting secondary analyses of their data from the University of New South Wales and the Australian National University. Figure 1 presents the CONSORT diagram of how participants were selected for Rasch analyses. All participants and informants provided written consent for their participation; this study was completed according to the guidelines of the contributing studies' institutions, which are based on internationally accepted ethical standards.

The OATS sample included 623 participants (600 twins and 23 siblings) aged above 65 years at the wave 1/baseline assessment. Of these, 468 completed the IQCODE-16 and were thus included in the current study. The OATS sample was aged from 65 to 92 ($M = 70.43$, $SD = 5.46$), and comprised 288 females (62%) and 180 males (39%).

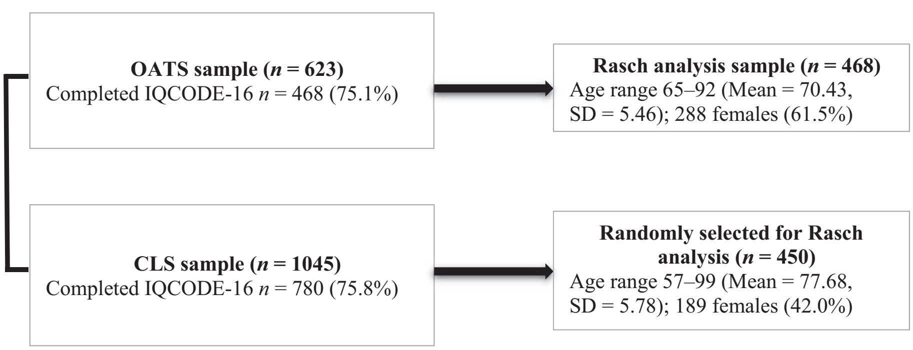

Figure 1 CONSORT diagram for participants selected for Rasch analysis of the 16-item Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE-16). CLS, Canberra Longitudinal Study; OATS, Older Australian Twins Study.

The original CLS sample included 1045 participants aged 57 or above at the wave 1/baseline assessment; of these 780 completed the IQCODE-16, and 450 were randomly drawn for Rasch analysis. This selection made the number of participants in this sample roughly similar to the OATS sample and satisfied the optimal sample size to minimize Type I and Type II errors of 250 to 500 for Rach analysis.[49,50] This selected CLS sample was aged 57–99 (M = 77.68, SD = 5.78), and comprised 189 females (42%) and 261 males (58%).

Moreover, we included the sample from the Sydney MAS that was used in Truong and colleagues' original study.[21] The baseline (wave 1) MAS sample included 1037 participants aged between 70 and 90. MAS participants were predominantly European (98%) and were recruited from the Eastern suburbs of Sydney, Australia, between 2005 and 2007.[13] The MAS sample selected for Rasch analysis in Truong and colleagues'[21] study comprised 400 participants.

We then randomly selected 150 participants from each of three original samples (OATS, CLS, and MAS) for invariance (DIF) testing across samples and to determine the overall model fit, as well as to produce a conversion table that is more representative of the general older Australian population. This random sample included participants aged 65–95 (M = 75.36, SD = 6.27), and comprised 254 females (56.4%) and 196 males (43.6%). Table S1 presents demographic details of participants in each sample used for the Rasch analysis. The results of chi-square and ANOVA tests showed that there were significant differences between age and sexes among three Rasch samples (*P*s < 0.001).

It should be noted that the full CLS sample (n = 468) was used for Rasch analysis because this sample size satisfied requirements of the Rasch model to reduce Type I and Type II error, which is between 250 and 500 cases. To ensure valid comparisons with the CLS sample and to satisfy Rasch model requirements, two comparable samples with n = 450 each were extracted using Simple Random Sampling (e.g. computer-generated randomization). The first comparison sample was comprised of 450 OATS participants and the second comparison sample was comprised of 450 participants randomly selected from all three original datasets (i.e. OATS, CLS, and MAS).

Measure

Informant-observed CCs were measured using the IQCODE-16,[18] which consists of 16 individual items that ask informants about how the participant's memory and cognitive function have changed (e.g. 'Using his/her intelligence to understand what's going on and to reason things through'). Each item is scored on a 5-point Likert-scale with response options ranging from 1 = 'much improved' to 5 = 'much worse', where 3 = 'no change'.

Data analyses

Descriptive statistics including mean, standard deviation (SD), Cronbach's alpha, and McDonald's omega for the IQCODE-16 across the four samples were computed using IBM SPSS v.27. Rasch analyses were conducted using the RUMM2030 software package[51] replicating the analytical solutions in Truong and colleagues,[21] which were based on the standardized criteria for the Rasch model fit as recommended elsewhere.[39,52]

Replications involved the use of unrestricted Partial Credit models[24] and examined whether the Rasch model fit solution in Truong and colleagues'[21] study was appropriate for all samples in the current study. The previous solution combined individual items into super-items (testlets) to reduce measurement error and improve the Rasch model fit.[21,32,53,54] Generally, the scale satisfies expectations of the Rasch model if it is not significantly different from the characteristics of an interval measure. This is reflected by nonsignificant item-trait interactions, no misfitting items, no local dependency, no DIF, and evidence of unidimensionality.[39,52] First, the overall Rasch model fit requires a nonsignificant chi-square index of the estimate of item-trait interaction ($P > 0.05$).[55] Second, no misfitting items can be identified, meaning fit residuals for individual items are in the range of ± 2.50.[29] Third, there can be no local dependency detected when observing the residual correlations between individual items, meaning values are below 0.20.[33] Fourth, no DIF due to personal factors (e.g. age, sex, sample) can be detected, which suggests the scale items work equally well across different groups of people.[41] Last, the achievement of unidimensionality is evidenced by a nonsignificant principal components analysis of the residuals and the equating *t*-test.[52] In addition, the Person Separation Index (PSI), the reliability coefficient used in Rasch analysis, is not a criterion of the Rasch model fit but it reflects how well the scale discriminates between individuals with different levels of the latent trait (e.g. CC severity). PSI values ranging from 0.70 to 0.80 indicate acceptable reliability and a PSI above 0.80 indicates good to excellent reliability.[51,53]

The best model fit for this study involved a solution that worked equally well across all included samples. When the best Rasch model fit was achieved, the person-item thresholds distribution for each sample was evaluated to examine how well items thresholds of the IQCODE-16 cover the sample's CC levels. A transformation table was then generated to transform the IQCODE-16 raw scores into interval-transformed scores to increase the precision of assessment.

RESULTS

The IQCODE-16 showed excellent internal consistency, with both Cronbach's alpha and McDonald's omega ranging from 0.93 to 0.96, which is consistent with previous reports.[5,19–21] Table 1 presents estimates of item-fit residuals for the initial Rasch analyses for each sample. As can be seen, most items with significant misfit to the Rasch model in the MAS sample from the Truong *et al*., study[21] (i.e. items 3, 10, 12, and 15) were consistently found to be misfitting in all samples in the present analyses, the exception being item 15, which was not misfitting in the random sample. Besides that, other items that misfit the Rasch model were observed across the OATS, CLS, and random samples. These misfitting items were 2, 5, 9, 13, and 16, while items 1 and 8 were only significantly misfit to the OATS sample. Table 2 displays the overall model fit estimates of the Rasch analyses for each IQCODE-16 items across all samples. All initial analyses (labelled A1) for the OATS, CLS, and random samples resulted in good to excellent reliability with PSIs ranging from 0.84 to 0.90. However, the overall fit to the Rasch model was not satisfactory as indicated by significant chi-square indexes (*P*s < 0.004), which reflect deviation of the scale from the Rasch model expectations across samples.

The overall model fit estimates using the Rasch model fit solution originally suggested by Truong and colleagues[21] across the four samples are also presented in Table 2 (analyses A2). This solution included four super-items: super-item 1 (items 2, 3, 13, and 14); super-item 2 (1 and 9); super-item 3 (6 and 7); and super-item 4 (10 and 11). These analyses demonstrated good to excellent reliability with PSIs ranging from 0.83 to 0.91 for the IQCODE-16, with an acceptable Rasch model fit as evident by nonsignificant chi-square indexes ($P > 0.05$) for the CLS and random samples. However, the OATS sample did not achieve Rasch model fit with this solution (chi-square index $P < 0.001$). The residual correlation matrix in the analysis for the OATS sample also showed local dependency between super-item 2 and super-item 4, and between super-item 2 and regular item 8 as reflected by significant residual correlations of 0.27 and 0.31, respectively.

To address these issues, minor modifications were made that involved combining locally dependent items in the OATS sample into three super-items: super-item 1 (items 2, 3, 13, and 14); super-item 2 (1 and 9); and super-item 3 (6, 7, 10, and 11). This new Rasch model provided not only a better fit for the OATS sample, but also for the CLS and random samples, and the MAS sample used in Truong and colleagues'[21] study (Table 2, analyses A3). This solution resolved item misfit and local dependency while achieving acceptable or strict unidimensionality, good reliability, and invariance across person factors such as participants' age and sex and the study sample. Figure 2 presents person-item threshold distributions from the analyses of the best fitting model for the samples in the current study as well as the MAS sample used in Truong and colleagues'[21] study. Interestingly, they show that thresholds of the IQCODE-16 scale satisfactorily cover CC levels with no significant ceiling or floor effects in each sample.

We generated a new conversion algorithm to transform raw IQCODE-16 scores into interval data using the best model fit analysis, allowing us to enhance precision of measurement. Table 3 presents a Rasch ordinal-to-interval conversion table which was developed based on person estimates for the IQCODE-16. Pearson's correlation between conversion scores in the current study and Truong and colleagues'[21] study indicated that the two sets of Rasch conversion scores were strongly correlated ($r = 0.998$).

Table 1 Rasch model fit statistics of item-fit residuals for the initial analyses (A1) of the IQCODE-16 individual items across four samples

Items	Item-fit residuals			
	MAS†	OATS	CLS	Random
1. Remembering things about family and friends (e.g. occupations, birthdays, addresses)	−1.66	−2.64*	−1.10	−2.61*
2. Remembering things that have happened recently	−2.52	−4.14*	−3.39*	−3.81*
3. Recalling conversations a few days later	−4.09*	−3.56*	−2.54*	−3.63*
4. Remembering his/her address and telephone number	0.48	−0.85	−1.30	−1.94
5. Remembering what day and month it is	−2.15	−2.54*	−3.32*	−2.74*
6. Remembering where things are usually kept	−1.45	−0.43	−2.11	−2.46
7. Remembering where to find things which have been put in a different place from usual	−2.23	−0.92	−2.02	−1.69
8. Knowing how to work familiar machines around the house	−2.07	−4.42*	−2.12	−2.32
9. Learning to use a new gadget or machine around the house	−2.35	−5.19*	−3.98*	−4.08*
10. Learning new things in general	−4.05*	−6.16*	−4.67*	−5.42*
11. Following a story in a book or on TV	−2.32	−2.04	−1.49	−2.31
12. Making decisions on everyday matters	−3.67*	−3.93*	−6.73*	−4.08*
13. Handling money for shopping	−0.51	−4.03*	−3.87*	−3.74*
14. Handling financial matters for example the pension, dealing with the bank	−1.84	−4.06*	−4.19*	−3.42*
15. Handling other everyday arithmetic problems (e.g. knowing how much food to buy, knowing how long between visits from family or friends)	−2.84*	−3.97*	−5.22*	−1.95
16. Using his/her intelligence to understand what's going on and to reason things through	−1.71	−5.78*	−4.35*	−3.16*

Abbreviations: IQCODE-16, 16-item Informant Questionnaire on Cognitive Decline in the Elderly; MAS, Memory and Ageing Study; OATS, Older Australian Twins Study; CLS, Canberra Longitudinal Study. * Significant misfit to the Rasch model. † Results for MAS were reproduced with permission from Truong and colleagues[21] Table 2.

Table 2 Summary of fit statistics for the Rasch analyses of the IQCODE-16 across samples

Sample	Analysis	Person mean		Goodness of fit		PSI	Significant *t*-tests (unidimensionality)	
		Value	SD	χ^2	*P*		%	Lower bound
MAS	A1†	1.41	2.26	92.23	0.16	0.92	5.5	3.4 (Acceptable)
	A2†	1.26	2.11	81.32	0.38	0.92	4.3	2.1 (Strict)
	A3‡	1.14	1.93	46.66	0.40	0.92	6.3	4.2 (Acceptable)
OATS	A1	0.02	1.58	191.51	<0.001	0.84	5.6	3.6 (Acceptable)
	A2	−0.08	1.52	93.13	<0.001	0.83	11.0	9.0 (Not acceptable)
	A3	−0.05	1.63	68.42	0.50	0.83	6.5	4.5 (Acceptable)
CLS	A1	0.76	1.69	138.94	0.003	0.90	4.9	2.8 (Strict)
	A2	1.04	1.82	30.99	0.06	0.91	8.1	6.0 (Not acceptable)
	A3	0.70	1.78	22.71	0.20	0.91	6.9	4.8 (Acceptable)
Random	A1	0.39	1.50	58.10	0.003	0.86	11.3	9.3 (Not acceptable)
	A2	0.30	1.65	92.33	0.16	0.85	8.1	6.1 (Not acceptable)
	A3	0.30	1.48	19.48	0.36	0.85	3.9	1.9 (Strict)

Abbreviations: A1, initial analysis; A2, replicating analysis; A3, Rasch model fit analysis; IQCODE-16, 16-item Informant Questionnaire on Cognitive Decline in the Elderly; MAS, Memory and Ageing Study; PSI, Person Separation Index; OATS, Older Australian Twins Study; CLS, Canberra Longitudinal Study. † Results for MAS were reproduced with permission from Truong and colleagues[21] Table 1. ‡ Analysis used MAS data with permission from Truong and colleagues.[21]

DISCUSSION

This study applied Rasch analysis to investigate psychometric properties of a widely used informant-reported CC measure, the IQCODE-16, and re-evaluated the generalizability of the Rasch conversion algorithms proposed in a recent study by Truong and colleagues.[21]

Good reliability of the IQCODE-16 was consistently obtained across analyses in this study, adding further empirical evidence supporting the robust psychometric properties of the scale. Interestingly, the Rasch solution suggested by Truong and colleagues[21] did not work equally well for all samples in the current

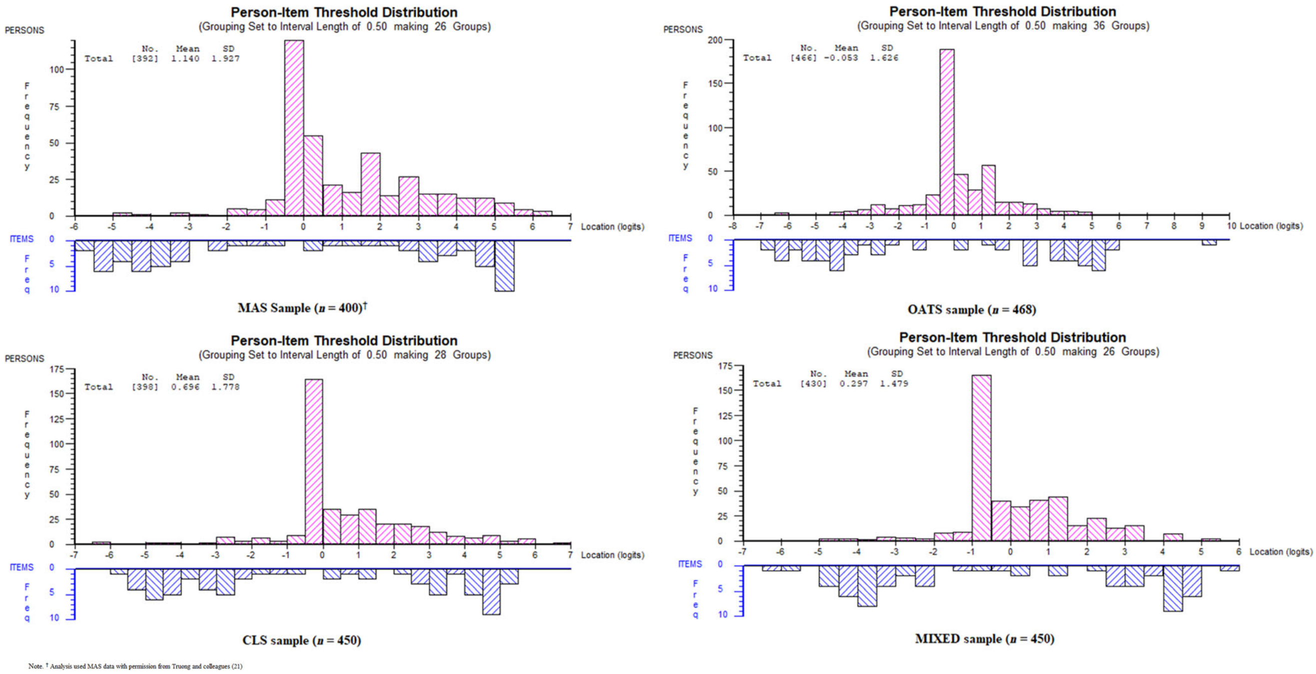

Figure 2 Person-item threshold distribution of the model fit analyses of the IQCODE-16 across four samples. CLS, Canberra Longitudinal Study; MAS, Memory and Ageing Study; OATS, Older Australian Twins Study.

study, potentially due to the previous study's biased sample. Therefore, an alternative solution was found in the current study, which involved reorganizing super-items of the IQCODE-16 to achieve the best Rasch model fit across all samples. This allowed us to generate ordinal-to-interval conversion algorithms to enhance the accuracy of the IQCODE-16 across more diverse samples with higher generalizability of assessment scores.

Evidence shows that the self-reported CCs can be significantly affected by individual person factors such as mood, personality, life events, and medications, which can also impact informants' reports. However, to the best of our knowledge, no interval-transformed algorithms have been previously established for self-reported CC assessments (e.g. MAC-Q). This is important as studies have shown that interval-transformed data can more accurately reflect real clinical changes (e.g. CC levels) compared to ordinal scales.[21,25,32] Moreover, using interval-transformed data can also decrease measurement errors associated with ordinal scale scores.[30] Therefore, the ordinal-to-interval transformation algorithms developed by this study contribute to the higher precision of the IQCODE-16 by lessening the influences of personal and affective factors on the informant's observation.

It should be noted that the Rasch solution fit in Truong and colleagues'[21] study was not replicated in this study, possibly because we used different DIF factors in the current study. Also, the differences between the original conversion algorithm generated by Truong and colleagues[21] and the one generated in the current study are marginal, meaning both conversion tables can be used for ordinal-to-interval transformations of the data. However, the current conversion table has a higher degree of generalizability across different samples and is more robust with regard to personal factors such as participants' age and sex. Moreover, the Rasch solution used in this study also worked well with the MAS sample used in Truong and colleagues'[21] study, even with different personal factors (i.e. informant age and sex), suggesting the conversion table produced in the current study is a superior option.

Each individual item of the original IQCODE-16 varies by its degree of difficulty, and hence contributes uniquely to the raw score, which should be considered when computing the overall score.[40] Traditional methods (e.g. classical test theory or generalizability theory) do not account for unequal contribution of items to the total score, while our

Table 3 Converting ordinal scores into interval-level scores for the IQCODE-16

Raw scores	Logits	Interval scores	Raw scores	Logits	Interval scores
1.00	−6.02	1.00	3.06	−0.10	2.79
1.06	−5.57	1.14	3.13	0.33	2.92
1.13	−5.28	1.22	3.19	0.73	3.04
1.19	−5.10	1.28	3.25	1.11	3.16
1.25	−4.96	1.32	3.31	1.46	3.26
1.31	−4.84	1.35	3.38	1.78	3.36
1.38	−4.74	1.39	3.44	2.07	3.45
1.44	−4.65	1.41	3.50	2.33	3.53
1.50	−4.56	1.44	3.56	2.56	3.60
1.56	−4.48	1.47	3.63	2.77	3.66
1.63	−4.39	1.49	3.69	2.97	3.72
1.69	−4.31	1.52	3.75	3.15	3.78
1.75	−4.23	1.54	3.81	3.32	3.83
1.81	−4.15	1.56	3.88	3.47	3.87
1.88	−4.07	1.59	3.94	3.61	3.92
1.94	−3.99	1.61	4.00	3.75	3.96
2.00	−3.90	1.64	4.06	3.87	3.99
2.06	−3.81	1.67	4.13	3.99	4.03
2.13	−3.71	1.70	4.19	4.11	4.06
2.19	−3.60	1.73	4.25	4.22	4.10
2.25	−3.48	1.77	4.31	4.32	4.13
2.31	−3.35	1.81	4.38	4.43	4.16
2.38	−3.21	1.85	4.44	4.54	4.19
2.44	−3.06	1.90	4.50	4.65	4.23
2.50	−2.89	1.95	4.56	4.77	4.27
2.56	−2.71	2.00	4.63	4.90	4.31
2.63	−2.50	2.06	4.69	5.06	4.35
2.69	−2.27	2.14	4.75	5.24	4.41
2.75	−2.00	2.22	4.81	5.48	4.48
2.81	−1.69	2.31	4.88	5.81	4.58
2.88	−1.33	2.42	4.94	6.34	4.74
2.94	−0.94	2.54	5.00	7.20	5.00
3.00	−0.52	2.66			

Abbreviation: IQCODE-16, 16-item Informant Questionnaire on Cognitive Decline in the Elderly.

interval-transformed scores generated by Rasch analyses can precisely estimate the unique contribution of each item to the overall assessment score, therefore contributing to higher accuracy of assessment. Studies suggest that Rasch interval-transformed data are more likely to reflect the CC levels of an individual accurately.[21,56] As such, using Rasch interval-transformed data is crucial because it decreases measurement error associated with raw ordinal scores.[57] Besides, such interval-level data are also suitable for conducting parametric statistics or statistical comparisons against other interval measurements (e.g. biomarkers, and electrophysiological and neuroimaging data) and may increase reliability and validity of the results because using interval data can avoid the violations of arithmetic assumptions inherent in ordinal data.[56]

The main strength of this study was the application of Rasch analysis methodology to the IQCODE-16 across several cohorts using an appropriate sample size that allowed us to minimize both Type I and Type II errors. Type I error occurs due to inflated chi-square statistics in RUMM2030 if the sample size exceeds 500 cases, while Type 2 error is common in smaller samples below 250 cases, as this limits the robustness of the item calibration.[49,50] In addition, this study is novel as, to date, there have been no replication studies using Rasch analysis to re-evaluate the psychometric properties of the 16-item version of the IQCODE. Moreover, this study found that the modified 16-item IQCODE works equally well across all samples and personal factors of the participants (i.e. participants' age and gender) which were not investigated in the previous study.

However, there are limitations which should be acknowledged. Although we considered older adults across three cohort studies, they were all Australian studies and may therefore not be representative of older adults in other regions, especially those in lower income countries. Therefore, this study should be replicated in different samples with different personal factors to investigate potential DIFs, for example, older adults from non-English speaking countries, or in low- and middle-income countries. It should also be noted that this study inclusively focused on reliability and internal validity because external validity of the IQCODE-16 is well established by other studies.[5,22,23]

In conclusion, the findings of this study demonstrated the reliability and internal validity of the informant-reported IQCODE-16 measure of CCs across older Australians. Our modification of the IQCODE-16, made using Rasch analyses, resolved local dependency issues and established scale invariance across different samples and personal factors. This allowed us to generate transformation tables to convert raw ordinal scores into interval-level data, which improves the precision of measurement. The interval-transformation table is more robust compared to a corresponding table generated in a previous study by our group. Clinicians and researchers can employ the IQCODE-16 in a variety of contexts with higher precision by using the conversion table published here, without needing any modification to the original IQCODE-16 administration format.

COMPLIANCE WITH ETHICAL STANDARDS

The study complied with the guidelines of the university ethics committee, which were internationally accepted ethical standards.

DISCLOSURE

The authors have no potential conflicts of interest to disclose.

ACKNOWLEDGMENT

We thank the participants and their informants for their time and generosity in contributing to this research. Open access publishing facilitated by The University of Waikato, as part of the Wiley - The University of Waikato agreement via the Council of Australian University Librarians.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

1 Crook TH, Feher EP, Larrabee GJ. Assessment of memory complaint in age-associated memory impairment: the MAC-Q. *Int Psychogeriatr* 1992; **4**: 165–176.
2 Jorm AF. The Informant Questionnaire on cognitive decline in the elderly (IQCODE): a review. *Int Psychogeriatr* 2004; **16**: 275–293.
3 Brodaty H, Pond D, Kemp NM *et al*. The GPCOG: a new screening test for dementia designed for general practice. *J Am Geriatr Soc* 2002; **50**: 530–534.
4 Petersen RC, smith GE, Waring sC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* 1999; **56**: 303–308.
5 Truong QC, Choo C, Numbers K *et al*. Applying generalizability theory to examine assessments of subjective cognitive complaints: whose reports should we rely on – participant versus informant? *Int Psychogeriatr* 2021: 1–11.
6 Numbers K, Crawford JD, Kochan NA, Draper B, Sachdev PS, Brodaty H. Participant and informant memory-specific cognitive complaints predict future decline and incident dementia: findings from the Sydney Memory and Ageing Study. *PloS One* 2020; **15**: e0232961.
7 Skoog I, Börjesson-Hanson A, Kern S *et al*. Decreasing prevalence of dementia in 85-year olds examined 22 years apart: the influence of education and stroke. *Sci Rep* 2017; **7**: 1–8.
8 Mitchell A, Beaumont H, Ferguson D, Yadegarfar M, Stubbs B. Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: meta-analysis. *Acta Psychiatr Scand* 2014; **130**: 439–451.
9 Nosheny RL, Amariglio R, Sikkes SA *et al*. The role of dyadic cognitive report and subjective cognitive decline in early ADRD clinical research and trials: current knowledge, gaps, and recommendations. *Alzheimer's Dement* 2022; **8**: e12357.
10 Caselli RJ, Chen K, Locke DE *et al*. Subjective cognitive decline: self and informant comparisons. *Alzheimers Dement* 2014; **10**: 93–98.
11 Rami L, Mollica MA, García-Sanchez C *et al*. The subjective cognitive decline questionnaire (SCD-Q): a validation study. *J Alzheimers Dis* 2014; **41**: 453–466.
12 Slavin MJ, Sachdev PS, Kochan NA *et al*. Predicting cognitive, functional, and diagnostic change over 4 years using baseline subjective cognitive complaints in the Sydney Memory and Ageing Study. *Am J Geriatr Psychiatry* 2015; **23**: 906–914.
13 Sachdev PS, Brodaty H, Reppermund S *et al*. The Sydney Memory and Ageing Study (MAS): methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70–90 years. *Int Psychogeriatr* 2010; **22**: 1248–1264.
14 Ponds RW, Jolles J. Memory complaints in elderly people: the role of memory abilities, metamemory, depression, and personality. *Educ Gerontol* 1996; **22**: 341–357.
15 Buckley RF, Ellis KA, Ames D *et al*. Phenomenological characterization of memory complaints in preclinical and prodromal Alzheimer's disease. *Neuropsychology* 2015; **29**: 571–581.

16 Reid LM, MacLullich AM. Subjective memory complaints and cognitive impairment in older people. *Dement Geriatr Cogn Disord* 2006; **22**: 471–485.
17 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association, 1994.
18 Jorm A. A short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): development and cross-validation. *Psychol Med* 1994; **24**: 145–153.
19 Harrison JK, Fearon P, Noel-Storr AH, McShane R, Stott DJ, Quinn TJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. *Cochrane Database Syst Rev* 2015; **3**: CD010772.
20 Phung TKT, Chaaya M, Asmar K *et al*. Performance of the 16-item informant questionnaire on cognitive decline for the elderly (IQCODE) in an Arabic-speaking older population. *Dement Geriatr Cogn Disord* 2015; **40**: 276–289.
21 Truong QC, Choo C, Numbers K *et al*. Enhancing precision of the 16-item Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE-16) using Rasch methodology. *Int Psychogeriatr* 2021: 1–11.
22 Park MH. Informant questionnaire on cognitive decline in the elderly (IQCODE) for classifying cognitive dysfunction as cognitively normal, mild cognitive impairment, and dementia. *Int Psychogeriatr* 2017; **29**: 1461–1467.
23 Perroco TR, Damin AE, Frota NA *et al*. Short IQCODE as a screening tool for MCI and dementia: preliminary results. *Dement Neuropsychol* 2008; **2**: 300–304.
24 Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982; **47**: 149–174.
25 Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009; **13**: iii, ix-x, 1–177.
26 Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. *Arch Phys Med Rehabil* 1989; **70**: 308–312.
27 Jamieson S. Likert scales: how to (ab) use them? *Med Educ* 2004; **38**: 1217–1218.
28 Rasch G. *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks pædagogiske Institut, 1960.
29 Lundgren-Nilsson A, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM™) revisited. *J Rehabil Med* 2011; **43**: 884–891.
30 Medvedev ON, Krägeloh CU, Titkova EA, Siegert RJ. Rasch analysis and ordinal-to-interval conversion tables for the Depression, Anxiety and Stress Scale. *J Health Psychol* 2020; **25**: 1374–1383.
31 Medvedev ON, Berk M, Dean OM *et al*. A novel way to quantify schizophrenia symptoms in clinical trials. *Eur J Clin Invest* 2020; **51**: e13398.
32 Merkin AG, Medvedev ON, Sachdev PS *et al*. New avenue for the geriatric depression scale: Rasch transformation enhances reliability of assessment. *J Affect Disord* 2020; **264**: 7–14.
33 Christensen KB, Kreiner S, Mesbah M. *Rasch Models in Health*. New York: Wiley Online Library, 2013.
34 Lord F, Novick M. In: Reading MA, ed. *Statistical Theories of Mental Test Scores*. California: Addison-Wesley Publishing Company, 1968.
35 Cronbach LJ, Rajaratnam N, Gleser GC. Theory of generalizability: a liberalization of reliability theory. *Br J Stat Psychol* 1963; **16**: 137–163.
36 Fox CM, Jones JA. Uses of Rasch modeling in counseling psychology research. *J Couns Psychol* 1998; **45**: 30–45.
37 Rasch G, On general laws and meaning of measurement in psychology. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. **4**, Berkeley: University of California Press, 1961; 321–333.
38 Medvedev ON, Titkova EA, Siegert RJ, Hwang YS, Krageloh CU. Evaluating short versions of the five facet mindfulness questionnaire using Rasch analysis. *Mind* 2018; **9**: 1411–1422.
39 Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care Res* 2007; **57**: 1358–1362.
40 Stucki G, Daltroy L, Katz JN, Johannesson M, Liang MH. Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts. *J Clin Epidemiol* 1996; **49**: 711–717.
41 Hagquist C, Andrich D. Is the sense of coherence-instrument applicable on adolescents? A latent trait analysis using Rasch-modelling. *Personal Individ Differ* 2004; **36**: 955–968.
42 Medvedev ON, Krägeloh CU. Rasch measurement model. In: *Handbook of Assessment in Mindfulness Research*. Switzerland: Springer, 2022; 1–18.
43 Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C. Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care* 2004; **42**: I25–I36.
44 Linacre JM. Estimation methods for Rasch measures. In: *Introduction to Rasch Measurement*. Minnesota: JAM Press, 2004; 25–48.
45 Tang WK, Wong E, Chan SS *et al*. The scoring scheme of the informant questionnaire on cognitive decline in the elderly needs revision: results of rasch analysis. *Dement Geriatr Cogn Disord* 2004; **18**: 250–256.
46 Sachdev PS, Lammel A, Trollor JN *et al*. A comprehensive neuropsychiatric study of elderly twins: the Older Australian Twins Study. *Twin Res Hum Genet* 2009; **12**: 573–582.
47 Christensen H, Mackinnon A, Jorm AF *et al*. The Canberra Longitudinal Study: design, aims, methodology, outcomes and recent empirical investigations. *Aging Neuropsychol Cogn* 2004; **11**: 169–195.
48 Jorm A, Christensen H, Korten A, Jacomb P, Henderson A. Informant ratings of cognitive decline in old age: validation against change on cognitive tests over 7 to 8 years. *Psychol Med* 2000; **30**: 981–985.
49 Azizan NH, Mahmud Z, Rambli A. Rasch rating scale item estimates using maximum likelihood approach: effects of sample size on the accuracy and bias of the estimates. *Int J Adv Sci Technol* 2020; **29**: 2526–2531.
50 Hagell P, Westergren A. Sample size and statistical conclusions from tests of fit to the Rasch model according to the Rasch Unidimensional Measurement Model (Rumm) program in health outcome measurement. *J Appl Meas* 2016; **17**: 416–431.
51 Andrich D, Sheridan B, Luo G. *RUMM 2030 (Beta Version for Windows) Perth*. Western Australia: RUMM Laboratory Pty Ltd., 2009.
52 Leung Y-Y, Png M-E, Conaghan P, Tennant A. A systematic literature review on the application of Rasch analysis in musculoskeletal disease—a special interest group report of OMERACT 11. *J Rheumatol* 2014; **41**: 159–164.
53 Medvedev ON, Turner-Stokes L, Ashford S, Siegert RJ. Rasch analysis of the UK functional assessment measure in patients with complex disability after stroke. *J Rehabil Med* 2018; **50**: 420–428.
54 Lundgren-Nilsson A, Jonsdottir IH, Ahlborg G, Tennant A. Construct validity of the psychological general well being index (PGWBI) in a sample of patients undergoing treatment for

stress-related exhaustion: a rasch analysis. *Health Qual Life Outcomes* 2013; **11**: 2.
55 Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002; **3**: 85–106.
56 Leung SO. A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *J Soc Serv Res* 2011; **37**: 412–421.
57 Medvedev ON, Siegert RJ, Kersten P, Krägeloh CU. Improving the precision of the Five Facet Mindfulness Questionnaire using a Rasch approach. *Mind* 2017; **8**: 995–1008.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's website: http://onlinelibrary.wiley.com/doi//suppinfo.

Table S1. Demographic details of participants included in three samples