International Psychogeriatrics (2024), 36:3, 166–176 © The Author(s), 2021. Published by Cambridge University Press on behalf of International Psychogeriatric Association.
doi:10.1017/S1041610221002568

Enhancing precision of the 16-item Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE-16) using Rasch methodology

Quoc Cuong Truong,[1,2] Carol Choo,[3] Katya Numbers,[4] Alexander G. Merkin,[5,6] Perminder S. Sachdev,[4] Valery L. Feigin,[5] Henry Brodaty,[4] Nicole A. Kochan,[4] and Oleg N. Medvedev[1]

[1]*School of Psychology, University of Waikato, Hamilton, New Zealand*
[2]*Faculty of Psychology, Vietnam National University Ho Chi Minh City, University of Social Sciences and Humanities, Ho Chi Minh City, Vietnam*
[3]*College of Healthcare Sciences, Division of Tropical Health and Medicine, James Cook University, Queensland, Australia*
[4]*Centre for Healthy Brain Ageing (CHeBA), University of New South Wales, Sydney, New South Wales, Australia*
[5]*National Institute for Stroke and Applied Neurosciences, Auckland University of Technology, Auckland, New Zealand*
[6]*Centre for Precise Psychiatry and Neurosciences, Kaufbeuren, Germany*

Abstract

Objective: This study aimed to investigate psychometric properties and enhance precision of the 16-item Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE-16) up to interval-level scale using Rasch methodology.

Design: Partial Credit Rasch model was applied to the IQCODE-16 scores using longitudinal data spanning 10 years of biennial follow-up.

Setting: Community-dwelling older adults aged 70–90 years and their informants, living in Sydney, Australia, participated in the longitudinal Sydney Memory and Ageing Study (MAS).

Participants: The sample included 400 participants of the MAS aged 70 years and older, 109 out of those were diagnosed with dementia 10 years after the baseline assessment.

Measurements: The IQCODE-16.

Results: Initial analysis indicated excellent reliability of the IQCODE-16, Person Separation Index (PSI) = 0.92, but there were four misfitting items and local dependency issues. Combining locally dependent items into four super-items resulted in the best Rasch model fit with no misfitting or locally dependent items, strict unidimensionality, strong reliability, and invariance across person factors such as participants' diagnosis and relationship to their informants, as well as informants' age and sex. This permitted the generation of conversion algorithms to transform ordinal scores into interval data to enhance precision of measurement.

Conclusions: The IQCODE-16 demonstrated strong reliability and satisfied expectations of the unidimensional Rasch model after minor modifications. Ordinal-to-interval transformation tables published here can be used to increase accuracy of the IQCODE-16 without altering its current format. These findings could contribute to enhancement of precision in assessing clinical conditions such as cognitive decline in older people.

Keywords: subjective cognitive complaints, measurement, Informant Questionnaire on Cognitive Decline in the Elderly, Rasch analysis, reliability

Introduction

Older adults frequently report subjective cognitive complaints (SCC), which may reflect self-estimations of changes in cognitive functions. SCC can be reported by an individual or by their close kin, friends or caregivers, referred to as informants (Brodaty *et al.*, 2002; Jorm *et al.*, 1991). SCC contribute to a diagnosis of Subjective Cognitive Decline (SCD), which are cognitive complaints in the absence of impaired cognitive performance (Centers for Disease Control and Prevention, 2019). SCD is thought to be a pre-mild cognitive impairment (MCI) stage of dementia (Jessen *et al.*, 2014). Therefore, SCC also contribute to screening and diagnosis of MCI.

Correspondence should be addressed to: Oleg N. Medvedev, School of Psychology, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand. Phone: + 64 7 837 9212. Email: oleg.medvedev@waikato.ac.nz. Received 03 Jul 2021; revision requested 05 Aug 2021; revised version received 25 Aug 2021; accepted 19 Sep 2021. First published online 19 November 2021.

Recent studies have shown that self- and informant-reported SCC, in the absence of impaired performance, predict steeper rates of cognitive decline and incident dementia 6 years prior to onset (Numbers *et al.*, 2020) and are related to the presence of Alzheimer's disease dementia biomarkers such as amyloid plaques in the brain, tau proteins in cerebral spinal fluid (Amariglio *et al.*, 2015; Sierra-Rio *et al.*, 2016), and atrophy and/or hypometabolism (Jessen *et al.*, 2014; Striepens *et al.*, 2010). Together, these data suggest that SCC may be reflect the earliest detectable stage of dementia (Skoog *et al.*, 2017). Moreover, SCC are quick and easy to capture and may provide insight into cognitive changes over and above those captured by more time-consuming and costly formal standardized neuropsychological testing (Numbers *et al.*, 2020).

More recently, researchers have begun examining the predictive utility of self- versus informant-reported SCC, and their association with cognitive performance and/or future decline. In the MAS study, SCC were measured over 10 years using two scales: the participant-reported MAC-Q and the informant-reported Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) (Sachdev *et al.*, 2010). Self-reported SCC are often highly correlated with an individual's mood, personality traits (under/over-complaining behaviors), life events, and medications (Buckley *et al.*, 2015a; Ponds and Jolles, 1996) and can be impacted by loss of insight that occurs at later stages of dementia (Buckley *et al.*, 2015b). Attention has turned toward exploring the utility of informant-reported SCC regarding the participant, with results suggesting informant reports may be better indicators of objective performance than self-reports. Indeed, previous studies using the same sample have shown that informant IQCODE scores are a more reliable approximation of actual cognitive performance than self-reported SCC (Slavin *et al.*, 2015; Truong *et al.*, 2021) and are more predictive of future cognitive decline and incident dementia (Numbers *et al.*, 2020). Further, confirmation of cognitive changes from an informant are now a key SCD-plus criteria – or a feature of SCC that increases the likelihood of preclinical AD – according to the Subjective Cognitive Decline Initiative (SCD-I) (Jessen *et al.*, 2014). Therefore, informant-reported SCC may be more accurate representations of cognitive decline as they are not subject to such biases. Indeed, in clinical contexts, informant SCC reports are reliably used as individuals begin losing insight into their cognitive changes over the progress of preclinical stages and the debilitating course of dementia (American Psychiatric Association, 1994).

The IQCODE (Jorm *et al.*, 1991) is a well-established psychometric informant-reported measure of SCC with 26 items. The short 16-item scale version of the IQCODE (IQCODE-16; Jorm, 1994) is a widely used SCC assessment instrument, which has demonstrated comparable reliability to the original version with Cronbach's alpha ranging from 0.93 to 0.97 (Harrison *et al.*, 2015; Jorm, 1994; Phung *et al.*, 2015). The IQCODE-16 is completed by informants who are caregivers, close kin, or friends of the individual and know them well enough to comment on their memory (Jorm, 1994). Studies have shown that the IQCODE-16 predicts incident dementia and can be used as a screening tool for dementia (Park, 2017; Perroco *et al.*, 2008). However, the differences between response options of an ordinal scale (e.g. 1 and 2 vs 2 and 3) may not reflect the same amount of clinical change compared to the interval scale, especially as individual items may contribute different amount of information about the latent trait to the overall assessment score (Hobart and Cano, 2009). Therefore, psychometric properties of the IQCODE-16 should be thoroughly examined in order to improve the precision of the instrument up to an interval measure by utilizing an advanced methodology such as Rasch analysis (Rasch, 1960; Tennant and Conaghan, 2007).

Rasch analysis is a powerful statistical method used to examine reliability and internal validity of psychometric instruments, as well as their specific psychometric properties such as functioning of individual items (Rasch, 1960; 1961) that has increasingly become a gold standard in clinical assessments (Hobart and Cano, 2009; Lundgren-Nilsson and Tennant, 2011). Rasch model is unidimensional and based on assumptions that the response to any specific item of a scale is determined by both an individual's ability and an item's difficulty (Rasch, 1960; 1961). While every ordinal item in an instrument usually has the same categorical values, the total scale score may be biased because each individual item has a different contribution to the overall latent trait (i.e. SCC levels) represented by items of a scale (Rasch, 1960). Moreover, the differences between response options of individual items reflect varying levels of clinically important change (Masters, 1982). Rasch analysis can reduce these biases by estimating precise thresholds between response categories of individual items and the unique contribution of each individual item to the overarching trait being measured (Stucki *et al.*, 1996). Such precise estimations are possible because most individuals get higher scores on the easy items, while only a few score highly on difficult items (Tennant and Conaghan, 2007). When data fit to the Rasch model, ordinal scores can be converted into interval-level data by accounting for item

https://doi.org/10.1017/S1041610221002568 Published online by Cambridge University Press

difficulty and person ability, hence increasing the precision of measurement that was demonstrated in a number of studies across different areas of clinical assessments at both the group and individual level (Norquist *et al.*, 2004).

Other statistical methods such as Classical Test Theory and its extension, Generalisability Theory (Cronbach *et al.*, 1963), are not able to estimate the contribution of each individual item of the scale to the overall latent trait because they do not differentiate between item difficulty based on individual ability (Fox and Jones, 1998). An item-person threshold distribution is another advanced feature of the Rasch analysis that can graphically show how the range of item difficulties covers range of persons' ability. This graphic is useful to detect potentially significant ceiling or floor effects.

However, no Rasch analysis has been conducted on the English scale version of the IQCODE-16 to further enhance its accuracy up to interval-level scale more suitable for parametric statistics and clinical evaluations. In fact, only one study to date has applied Rasch analysis to investigate the psychometric properties of the 26-item Chinese version of IQCODE in a Chinese sample (Tang *et al.*, 2004). Although this study concluded that 4 out of the 26 items had statistically inadequate fit and should be removed from the scale, the authors did not modify the scale to enhance its psychometric properties. These four misfitting items were already excluded in the 16-item English version of the IQCODE, which was developed earlier (Jorm, 1994).

Nonetheless, Tang *et al.*'s (2004) study findings are not applicable to the English version of the IQCODE and may not be generalizable to older adults with and without cognitive impairment as the sample was selectively comprised of Chinese stroke survivors and their informants. Moreover, Person Separation Index (PSI) was not reported in this study, even though it is an essential reliability estimate in Rasch analysis (Tennant and Conaghan, 2007). PSI is considered a reliable alternative to sensitivity analysis because it reflects the IQCODE's ability to distinguish between individual SCC levels (Fisher, 1992). Finally, Tang *et al.* (2004) did not generate converging tables to transform raw scores into interval-level data. Such converging tables have been recommended in the recent reporting guidelines as essential for Rasch analytic studies (Leung *et al.*, 2014) as they allow clinicians and researchers to transform raw scores into interval-level data to achieve higher precision of the scale without modifying the scale's original response format (Merkin *et al.*, 2020).

Given no study to date has applied Rash analysis to the English version of the IQCODE-16, and several shortcomings of the Tang *et al.* (2004) study's methodology, the aim of the current study was to apply Rasch methodology to investigate and enhance psychometric properties of the 16-item IQCODE. This is also important because cross-sectional SCC provided by a patient may be different from informant-reported scores that are considered more reliable. We have also considered that the longitudinal (10 years) informant-based history captured by the IQCODE may have impact on assessment scores and included two independent samples from the first and the last wave (10 years apart) in the analysis to address this issue. A secondary aim was to produce convention tables that can be used to convert raw scores into interval-level data to enhance the precision of the scale, if an acceptable fit to the Rasch model is achieved.

Method

Participants

Participants were drawn from the Sydney Memory and Ageing Study (MAS), a longitudinal study of cognitive aging that included a baseline sample of 1037 community-dwelling older adults aged between 70 and 90 years, without dementia. MAS participants were recruited from the Eastern suburbs of Sydney, Australia, between 2005 and 2007 (Sachdev *et al.*, 2010). The vast majority of participants identified as European (98%), with 1% identifying as Asian, and 1.1% as other or not revealed. Most participants (97.3%) had informants, who were a close friend or family member that knew the participant sufficiently well to answer questions relating to the participant's memory, thinking, and daily functioning. Informants were required to have at least 1 h of contact with the participant per week; on average they had 8.3 h of weekly contact. Participants and informants were interviewed biennially from wave 1 (baseline) to wave 6 (10-year follow-up). All participants and informants provided written consent to participate in this study, which was approved by the University of New South Wales Human Ethics Review Committee (HC 05037, 09382, 14327).

Figure 1 presents the consort diagram of how participants were selected for Rasch analysis. To control for differential item functioning (DIF) between participants who went on to be diagnosed as normal versus dementia at the latest wave, we included all IQCODE-16 reports of participants who were diagnosed with dementia at wave 6 ($n = 109$). Studies indicated that the appropriate sample size for Rasch analysis is between 250 and 500 participants, because it allows a researcher to minimize Type I and Type II errors with questionnaires consisting of 15–20 items (Azizan *et al.*, 2020;

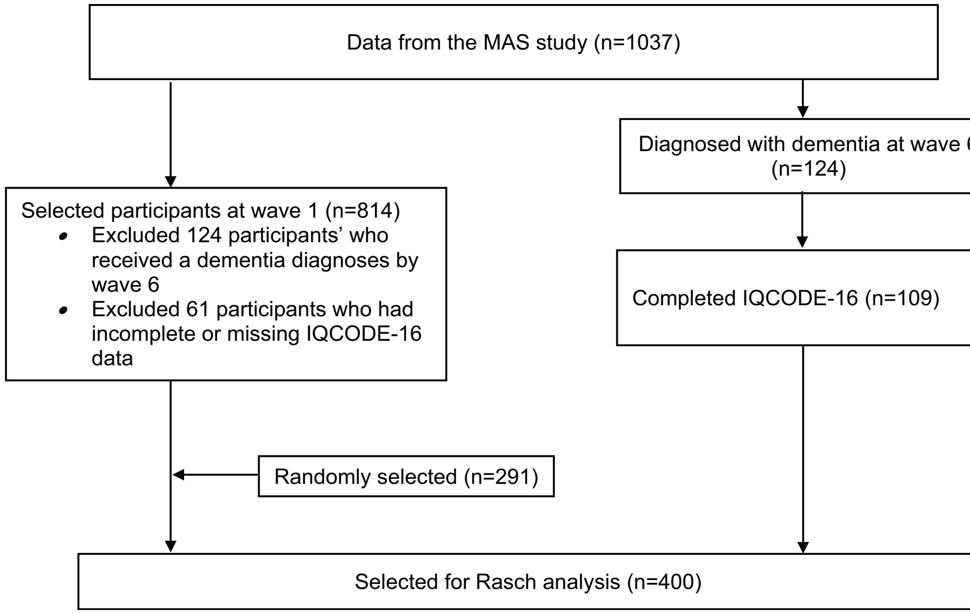

Figure 1. CONSORT flow diagram of participants selected for Rasch analysis of the IQCODE-16.

Hagell and Westergren, 2016). Therefore, to ensure unique participants in the sample and an appropriate sample size of 400 participants for Rasch analysis, 291 participants were then randomly selected at wave 1, after excluding data from participants who were diagnosed with dementia at wave 6. Informants in the selected sample were 121 males (30.3%) and 277 females (69.3%). Informant ages ranged from 24 to 95 years, with a mean age of 62.02 years (standard deviation, SD = 14.48). Ninety-eight informants were participants' spouses, 125 were participants' children, 2 were grandchildren, 7 were siblings, 16 were other relatives (e.g. niece), 64 were close friends, and 88 were "other." Missing data in the extracted sample comprised less than 0.01% and were completely at random.

Measure

The IQCODE-16 (Jorm, 1994) consists of 16 items that assess informant-reported SCC levels of participants. To complete the IQCODE-16, informants are asked questions about how the participants' memory and cognitive function have changed, regardless of their (participants') premorbid intelligence or education (Park, 2017). Items are scored on a 5-point Likert scale with options ranging from 1 = "much improved" to 5 = "much worse." For example, item 1 is "Remembering things about family and friends (e.g. occupations, birthdays, addresses)."

Data analyses

IBM SPSS v.27 was used to compute descriptive statistics including means, SD, and Cronbach's alpha for the IQCODE-16. RUMM2030 software package (Andrich *et al.*, 2009) was used to conduct Rasch analyses by following the recommendations of Tennant and Conaghan (2007) and standardized criteria to evaluate the Rasch model fit as described elsewhere (Leung *et al.*, 2014). As individual IQCODE-16 items were polytomous, a likelihood-ratio test was conducted to identify an appropriate polytomous Rasch model for analyses, the partial credit model (Masters, 1982) or the rating scale model (Andrich, 1978). While the rating scale model implies that response categories across all individual scale items have the same rating scale structure, the partial credit model is unrestricted and assumes that each individual item has its own unique response options structure (Linacre, 2000). The likelihood-ratio test examines similarity between thresholds of individual items. If threshold distances are significantly different across individual items, the partial credit model should be used (Lundgren-Nilsson and Tennant, 2011). Otherwise, the rating scale model would be suitable (Tennant and Conaghan, 2007).

Rasch analysis was iterative and continued until the best model fit was achieved. The overall model fit requires the estimate of item-trait interaction to be not significant, which is reflected by chi-square index ($p > 0.05$). The fit residuals for individual items were evaluated to detect item misfit (i.e. item fit residuals should be between −2.50 and +2.50). The residual correlations between individual items were then examined and those that had values above 0.20 were considered as indicative of local dependency (Christensen *et al.*, 2013). DIF due to relevant individual characteristics (i.e.

https://doi.org/10.1017/S1041610221002568 Published online by Cambridge University Press

Table 1. Summary of fit statistics for the initial, second, and third Rasch analyses of the IQCODE-16 (n = 400)

ANALYSES	PERSON MEAN		GOODNESS OF FIT		PSI	SIGNIFICANT *T*-TESTS (UNIDIMENSIONALITY)	
	VALUE/SD		χ2 (df)	*P*		%	LOWER BOUND
Initial (A1)	1.41	2.26	92.23(80)	0.16	0.92	5.5	3.4 (YES)
Second (A2)	1.36	2.22	85.70(86)	0.49	0.92	6.8	4.6 (YES)
Third (A3)	1.26	2.11	81.32 (78)	0.38	0.92	4.3	2.1 (STRICT)

personal factors) were also tested to identify whether all items were invariant across different groups (e.g. age, sex). Generally, the scale meets expectations of the Rasch model if there are no significant interactions between items and the latent trait, no misfitting items, no local dependency and/or DIF, and unidimensionality is evident (Leung *et al.*, 2014). Unidimensionality was examined using the principal component analysis of the residuals and paired *t*-test following the method developed by Smith (2002). In addition, PSI was used to evaluate reliability in Rasch analysis, which is not the Rasch model fit criteria but reflects how well the scale discriminates between individuals with different levels of the latent trait (e.g. SCC). PSI is interpreted somewhat similar to Cronbach's alpha with values above 0.70 indicating acceptable reliability for group assessments and 0.80 and higher for individual assessments (Medvedev *et al.*, 2018).

In this study, super-items were created by combining locally dependent items together to improve the Rasch model fit (Lundgren-Nilsson *et al.*, 2013; Medvedev *et al.*, 2018). When the Rasch model fit was satisfactory, the person-item thresholds distribution was examined showing how well items thresholds of the IQCODE-16 cover SCC levels of the sample. Lastly, the transformation table was produced to convert raw scores into interval-level data to increase the precision of assessment. Statistical significance was estimated using the conventional cut-off point of p-value > 0.05.

Results

The IQCODE-16 ($M = 52.45$, $SD = 7.98$) showed satisfactory internal reliability, with Cronbach's alpha value of 0.96, which is consistent with previous validation reports (Harrison *et al.*, 2015; Phung *et al.*, 2015).

A likelihood-ratio test showed significant differences between thresholds across individual IQCODE-16 items, $\chi^2(44) = 150.57$, $p < 0.001$, which means that the unrestricted partial credit model was more appropriate to use for the data of this study. Table 1 displays the overall model fit estimates of the initial, second, and third Rasch analyses of the IQCODE-16. As can be seen, initial analysis (A1) indicated excellent reliability PSI = 0.92 for the full IQCODE-16 and the overall fit to the Rasch model was acceptable as indicated by nonsignificant chi-square, $\chi^2(80) = 92.23$, $p = 0.16$. Table 2 presents estimates of the initial Rasch analysis for individual items including item location, fit residual, and chi-square values. There are four items (i.e. item 3, 10, 12, and 15) with significant misfit to the model.

The residual correlation matrix was also examined and showed that local dependency between several items was above 0.20. For example, the residual correlation between item 9 and 10 was 0.52, and between item 1 and 2, and 2 and 3, were both 0.32. Combining locally dependent items into super-items can reduce measurement error as well as improve individual items and the overall model fit (Lundgren-Nilsson *et al.*, 2013; Medvedev *et al.*, 2018; Merkin *et al.*, 2020). Therefore, five super-items comprised of locally dependent items were created: super-item 1 (Items 1&9); super-item 2 (2&13); super-item 3 (3&14); super-item 4 (6&7); and super-item 5 (10&11). This improved the overall model fit (see Table 1, analysis A2) with no change of reliability and acceptable unidimensionality (4.6% of significant *t*-tests). However, the residual correlation between super-item 1 and super-item 3 was still above 0.20, which indicated local dependency. To address this issue, the third analysis (A3) was conducted with four super-items (i.e. super-item 1: items 2, 3, 13, & 14; super-item 2: items 1 & 9; super-item 3: items 6 & 7, and super-item 4: items 10 & 11). This analysis resulted in strict unidimensionality and excellent reliability (see Table 1). Besides that, examination of the residual correlations indicated no local dependency and no DIF amongst items/super-items by personal factors. Therefore, this analysis (A3) achieved the best Rasch model fit for the IQCODE in the current sample.

Figure 2 presents person-item threshold distribution from the analysis of the best model fit (analysis A3) for the IQCODE-16. It shows that the scale's thresholds satisfactory cover SCC levels of the

https://doi.org/10.1017/S1041610221002568 Published online by Cambridge University Press

Table 2. Rasch model fit statistics including item locations, fit residuals, and chi-square for the initial analysis of the IQCODE-16 individual items

ITEM	ITEM LOCATION	ITEM-FIT RESIDUAL	CHI-SQUARE
1. Remembering things about family and friends (e.g. occupations, birthdays, addresses)	– 0.01	– 1.66	5.58
2. Remembering things that have happened recently	– 0.03	– 2.52	6.72
3. Recalling conversations a few days later	– 0.91	– 4.09*	6.02
4. Remembering his/her address and telephone number	0.59	0.48	8.87
5. Remembering what day and month it is	0.15	– 2.15	3.48
6. Remembering where things are usually kept	– 0.61	– 1.45	1.11
7. Remembering where to find things which have been put in a different place from usual	– 1.04	– 2.23	20.47
8. Knowing how to work familiar machines around the house	0.04	– 2.07	3.88
9. Learning to use a new gadget or machine around the house	– 0.34	– 2.35	7.52
10. Learning new things in general	– 0.30	– 4.05*	4.60
11. Following a story in a book or on TV	0.13	– 2.32	1.38
12. Making decisions on everyday matters	<0.01	– 3.67*	2.68
13. Handling money for shopping	0.55	– 0.51	7.43
14. Handling financial matters, for example, the pension, dealing with the bank	0.76	– 1.84	2.98
15. Handling other everyday arithmetic problems (e.g. knowing how much food to buy, knowing how long between visits from family or friends)	0.28	– 2.84*	7.05
16. Using his/her intelligence to understand what's going on and to reason things through	0.72	– 1.71	2.43

* Significant misfit to the Rasch model.

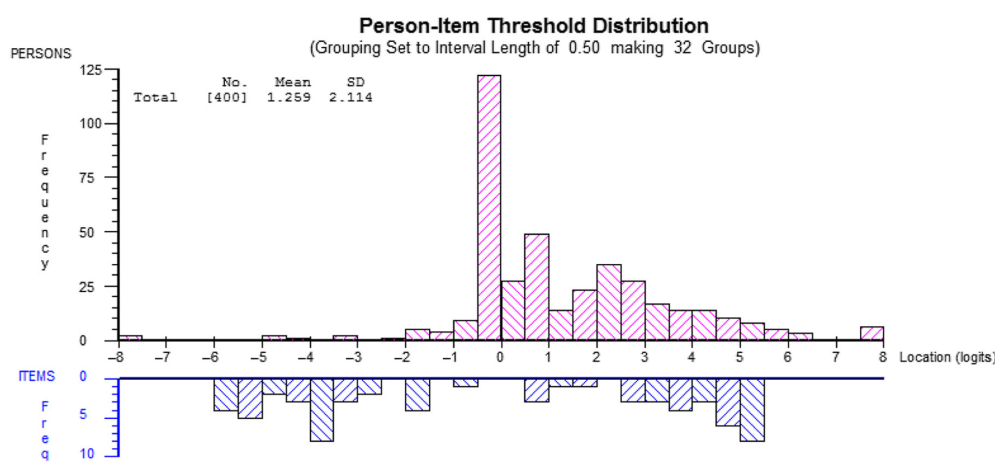

Figure 2. Person-item threshold distribution of the best model fit analysis (A3) of the IQCODE-16.

sample and there are no significant ceiling or floor effects. The sample mean is higher than the item means, reflecting the overall higher SCC levels of the sample due to inclusion of the subsample of participants with dementia diagnosis, though the mode was just below zero mark.

The third analysis (A3) for the IQCODE-16 demonstrated the best Rasch model fit that permitted conversion of raw ordinal IQCODE-16 scores into interval-level data. Table 3 displays Rasch ordinal-to-interval transformation table developed for the IQCODE-16 based on person estimates of the model.

Paired sample t-tests were used to examine the difference between IQCODE-16 raw scores and interval transform scores at wave 6 because at this wave more participants were diagnosed with dementia (n = 124). The results revealed that for the full sample, ordinal raw scores ($M = 53.22$, $SD = 7.79$) were significantly lower compared to interval-level Rasch scores ($M = 53.53$, $SD = 8.24$) as evidenced by the test statistics ($t(337) = -3.21$, $p = 0.001$). Similarly, ordinal raw scores ($M = 61.12$, $SD = 10.13$) were significantly lower compared to interval-transformed scores ($M = 61.71$, $SD = 9.53$) in the subsample consisted of those who were

https://doi.org/10.1017/S1041610221002568 Published online by Cambridge University Press

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

RAW SCORES	LOGITS	INTERVAL SCORES	RAW SCORES	LOGITS	INTERVAL SCORES
16	−7.65	16.00	49	0.15	48.58
17	−6.93	19.01	50	0.58	50.38
18	−6.45	21.01	51	0.99	52.09
19	−6.13	22.34	52	1.38	53.69
20	−5.89	23.37	53	1.73	55.18
21	−5.68	24.24	54	2.06	56.53
22	−5.49	25.01	55	2.35	57.76
23	−5.32	25.72	56	2.61	58.86
24	−5.17	26.38	57	2.85	59.86
25	−5.02	27.01	58	3.07	60.76
26	−4.87	27.60	59	3.27	61.59
27	−4.74	28.17	60	3.45	62.36
28	−4.61	28.72	61	3.62	63.07
29	−4.48	29.25	62	3.78	63.74
30	−4.35	29.78	63	3.93	64.36
31	−4.23	30.30	64	4.07	64.96
32	−4.10	30.83	65	4.21	65.53
33	−3.97	31.37	66	4.34	66.07
34	−3.84	31.93	67	4.47	66.61
35	−3.70	32.51	68	4.60	67.13
36	−3.55	33.12	69	4.72	67.65
37	−3.40	33.76	70	4.85	68.18
38	−3.23	34.45	71	4.97	68.71
39	−3.06	35.18	72	5.11	69.26
40	−2.87	35.98	73	5.25	69.85
41	−2.66	36.86	74	5.40	70.50
42	−2.42	37.83	75	5.57	71.22
43	−2.16	38.92	76	5.78	72.07
44	−1.86	40.16	77	6.03	73.13
45	−1.53	41.56	78	6.37	74.55
46	−1.15	43.14	79	6.89	76.71
47	−0.74	44.87	80	7.68	80.00
48	−0.29	46.72			

diagnosed with dementia at wave 6 ($t(81) = -2.52$, $p = 0.01$) but not in the subsample of those who were not dementia diagnosed ($p > 0.05$).

Discussion

The IQCODE-16 is one of the most widely used tools to assess SCC levels in older persons. This study applied Rasch analysis to investigate psychometric properties of the IQCODE-16 and derived ordinal-to-interval conversion tables to enhance precision of the instrument. Achieving excellent reliability (PSI = 0.92) in this study provides further empirical evidence to support the robust psychometric properties of the 16-item IQCODE and suitability of the modified scale for both individual and group assessment. Other findings of the current study demonstrated that the IQCODE-16, reorganized into super-items, achieved the best fit to the unidimensional Rasch model, which then permitted us to improve precision of the measure using the ordinal-to-interval conversion algorithms presented in Table 3. Such transformation is important because individual items of ordinal scales such as the IQCODE-16 have varying degrees of difficulty and thus, each item contributes uniquely to the total score, which should be accounted for (Stucki *et al.*, 1996). As such, using Rasch transformation tables decreases measurement error associated with ordinal scales scores (Medvedev *et al.*, 2017). We have also demonstrated that the ordinal scale bias was statistically significant in the current study as evident by paired *t*-tests comparing raw scores and interval-transformed scores converted to the same scale range in the total sample and dementia-diagnosed subsample. The ordinal IQCODE raw scores were significantly lower than interval-level scores in the sample of participants who were diagnosed with dementia, though the effect size of this difference was relatively small. However, this illustrates that using IQCODE-16

https://doi.org/10.1017/S1041610221002568 Published online by Cambridge University Press

ordinal scores may impact on the results if conducting parametric statistical tests. An extensive literature on Rasch methodology suggests that the interval-level scores derived from Rasch analyses may more accurately reflect an individual's level of SCC and could thus be used to conduct parametric statistics without violating their arithmetic assumptions (Leung, 2011).

Interval-level data of the IQCODE-16 can also be used for valid statistical comparisons with other interval measures like electrophysiological and neuroimaging data and biomarkers, and earlier collected data can be reanalyzed/replicated to increase reliability and validity of the results. Further, transforming the IQCODE-16 ordinal scores into interval data is user-friendly and does not require expertise in statistics, meaning that researchers can easily determine an individual's interval score based on their raw IQCODE-16 score simply by referring to the data in Table 3. Moreover, an ordinal-to-interval conversion Syntax file in IBM SPSS format is available as a Supplemental file to simplify the conversion for researchers working with large datasets.

Ordinal-to-interval transformation tables also have important clinical implications as our study offers a parsimonious tool to assist clinicians in determining presence and severity of SCC. Although the evidence supported reliability and validity of the original IQCODE scores in the clinical context, these scores constitute an ordinal measure. For example, if a participant A has an IQCODE score of 32, they are at less risk to have cognitive impairment (e.g. dementia) compared to participant B who scores 59. Assume, then, that both participants A and B engage in the therapeutic intervention, and participant A's score reduces to 22 and participant B's score reduces to 49. We can see that both participants have their scores decreased by 10 points on the IQCODE ordinal scale. When using the Rasch interval transformed scores, participant A's score has decreased by 5.82 points, while participant B's scores has decreased by 13.01 points. This shows that participant B's reduction in SCC level is more considerable compared to the participant A, meaning that one can more accurately detect and evaluate real change when using Rasch measurement with interval-transformed data. When using the IQCODE to identify who does or does not meet screening criteria (e.g. cognitive decline), clinicians do not need to transform the ordinal scores because they work well for this purpose. The transformed scores should be used in clinical context where the IQCODE is used both for screening and outcome monitoring, but it should be noted that it is also important to convert the criterion in the interval metrics for relative evaluation.

Strengths

The main strength of this study was the application of modern and robust Rasch methodology to an adequate sample size within an optimal range ($250 < n \leq 500$) that permitted minimizing Type I and Type II errors with questionnaires consisting of 15–20 items (Azizan *et al.*, 2020; Hagell and Westergren, 2016). Besides that, this study is also novel in several ways. First, to date, there have been no studies investigating the psychometric properties of the 16-item English version of the IQCODE using Rasch methodology. Combining locally dependent items into four super-items resolved local dependency problems that can produce spurious correlations affecting scale accuracy and the Rasch model fit (Medvedev *et al.*, 2018). Appropriateness of using super-items in the current study was further supported by strict unidimensionality and invariance of the modified 16-item IQCODE, because research has shown that multidimensional scales cannot generate super-items with an adequate Rasch model fit (Mitchell-Parker *et al.*, 2018). In addition, the achievement of strict unidimensionality and invariance across all personal factors of the modified IQCODE-16 also indicated that the modified 16-item IQCODE works equally well for both healthy older persons and those with a diagnosis of dementia and is not impacted by the informant's gender, age, or relationship to the individual. It should be noted that the IQCODE-16's modifications were implemented internally and work if the ordinal-to-interval conversion tables are applied, which does not require modifications of the original administration format of the scale.

Limitations

Our study is not without some limitations, which should be acknowledged. Data used for the current study's analyses are highly homogenous and not representative of all older adults, given MAS study participants and informants were recruited from an affluent area of Sydney, Australia, with a predominantly White European ethnic group. Clinical research from culturally diverse data provides support for use of the IQCODE-16 to detect early stages of dementia, as there may be cultural variations across a range of health issues (Choo *et al.*, 2017). Additionally, it is worth noting that there was unbalance between informant genders, 693 (68.7%) and 277 (69.3%) females in the original sample and in the sample selected for Rasch analysis, respectively. Although Rasch methodology tends to be robust and less affected by sampling bias compared to other methods (Hobart and Cano, 2009), the results of this study should be replicated in a more diverse sample to investigate potential DIF across sample groups unrepresented in the current study. If DIF is found for a specific group (e.g. other English-speaking countries, less affluent groups, or samples that are more balanced with respect to informant genders or reflect other ethnic groups), an additional conversion table could be produced for such group to permit valid score comparisons across a wider population.

Conclusion

In conclusion, the findings of this study indicate that the IQCODE-16 is a reliable and valid assessment tool for measuring SCC among older adults. Our adjustments of the IQCODE-16, made by Rasch analyses, resolved local dependency issues permitting transformation of raw scores into interval-level data, which improve the precision of measurement. The interval-transformed data table allows both clinicians and researchers to apply this sound psychometric measurement in a variety of contexts with higher precision, without needing any modification to the original IQCODE-16 administration format.

Acknowledgements

The Sydney Memory and Ageing Study has been funded by three National Health & Medical Research Council (NHMRC) Program Grants (ID No. ID350833, ID568969, and APP1093083). We thank the participants and their informants for their time and generosity in contributing to this research. We also acknowledge the MAS research team: https://cheba.unsw.edu.au/research-projects/sydney-memory-and-ageing-study.

Conflicts of interest

None.

Description of author roles

QT led and designed the study, conducted data analyses, and wrote the manuscript. CC supervised the study and edited the manuscript. KN coordinated data management, collaborated with the study design, and edited the manuscript. AM and VF collaborated with designing of the study and editing of the manuscript. PS supervised data collection, collaborated with designing of the study and editing of the manuscript. NK supervised data collection and edited the manuscript. HB sourced funding, supervised data collection, collaborated with editing of the manuscript. OM supervised the study and data analyses and edited the manuscript.

Ethical standards

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

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1041610221002568

References

Amariglio, R. E. *et al.* (2015). Tracking early decline in cognitive function in older individuals at risk for Alzheimer disease dementia: the Alzheimer's Disease Cooperative Study Cognitive Function Instrument. *JAMA Neurology*, 72, 446–454.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–573.

Andrich, D., Sheridan, B. and Luo, G. (2009). *RUMM, 2030 (Beta Version for Windows) Perth*. Western Australia: RUMM Laboratory Pty Ltd.

Azizan, N. H., Mahmud, Z. and Rambli, A. (2020). Rasch rating scale item estimates using maximum likelihood approach: effects of sample size on the accuracy and bias of the estimates. *International Journal of Advanced Science and Technology*, 24(4), 2526–2531.

Brodaty, H. *et al.* (2002). The GPCOG: a new screening test for dementia designed for general practice. *Journal of the American Geriatrics Society*, 50, 530–534.

Buckley, R. F. *et al.* (2015a). Phenomenological characterization of memory complaints in preclinical and prodromal Alzheimer's disease. *Neuropsychology*, 29, 571–581.

Buckley, R. F. *et al.* (2015b). Self and informant memory concerns align in healthy memory complainers and in early stages of mild cognitive impairment but separate with increasing cognitive impairment. *Age and Ageing*, 44, 1012–1019.

Choo, C. C., Harris, K. M., Chew, P. K. and Ho, R. C. (2017). Does ethnicity matter in risk and protective factors for suicide attempts and suicide lethality? *PLOS ONE*, 12, e0175752.

Christensen, K. B., Kreiner, S. and Mesbah, M. (2013). *Rasch Models in Health*. Hoboken, NJ: John Wiley and Sons.

Centers for Disease Control and Prevention (2019). *Subjective Cognitive Decline—A Public Health Issue*. Retrieved on 24 of August 2021 from https://www.cdc.gov/aging/data/subjective-cognitive-decline-brief.html

Cronbach, L. J., Rajaratnam, N. and Gleser, G. C. (1963). Theory of generalizability: a liberalization of reliability theory. *British Journal of Statistical Psychology*, 16, 137–163.

Fisher, W. P. Jr 1992. Reliability statistics. *Rasch Measurement Transactions*, 6, 238. http://www.rasch.org/rmt/rmt63i.htm.

Fox, C. M. and Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45, 30–45.

Hagell, P. and Westergren, A. (2016). 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. *Journal of Applied Measurement*, 17, 416–431.

Harrison, J. K., Fearon, P., Noel-Storr, A. H., McShane, R., Stott, D. J. and Quinn, T. J. (2015). Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a secondary care setting. *Cochrane Database of Systematic Reviews*, 12, 203.

Hobart, J. and Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technology Assessment*, 13(12).

Jessen, F. *et al.* (2014). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's and Dementia*, 10, 844–852.

Jorm, A. (1994). A short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): development and cross-validation. *Psychological Medicine*, 24, 145–153.

Jorm, A., Scott, R., Cullen, J. and MacKinnon, A. J. (1991). Performance of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) as a screening test for dementia. *Psychological Medicine*, 21, 785–790.

Leung, S. O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research* 37(4):412–421.

Leung, Y. Y., Png, M. E., Conaghan, P. and Tennant, A. (2014). A systematic literature review on the application of Rasch analysis in musculoskeletal disease—A special interest group report of OMERACT 11. *The Journal of Rheumatology*, 41, 159–164.

Linacre, J. M. (2000). Comparing "partial credit" and "rating scale" models. *Rasch Measurement Transactions*, 14, 768.

Lundgren-Nilsson, A., Jonsdottir, I. H., Ahlborg, G. and Tennant, A. (2013). 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 and Quality of Life Outcomes*, 11, 2.

Lundgren-Nilsson, A. and Tennant, A. (2011). Past and present issues in Rasch analysis: the functional independence measure (FIMTM) revisited. *Journal of Rehabilitation Medicine*, 43, 884–891.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.

Medvedev, O. N., Siegert, R. J., Kersten, P. and Krägeloh, C. U. (2017). Improving the precision of the Five Facet Mindfulness Questionnaire using a Rasch approach. *Mindfulness*, 8, 995–1008.

Medvedev, O. N., Turner-Stokes, L., Ashford, S. and Siegert, R. J. (2018). Rasch analysis of the UK Functional Assessment Measure in patients with complex disability after stroke. *Journal of Rehabilitation Medicine*, 50, 420–428.

Merkin, A. G. *et al.* (2020). New avenue for the geriatric depression scale: Rasch transformation enhances reliability of assessment. *Journal of Affective Disorders*, 264, 7–14.

Mitchell-Parker, K., Medvedev, O. N., Krägeloh, C. U. and Siegert, R. J. (2018). Rasch analysis of the frost multidimensional perfectionism scale. *Australian Journal of Psychology*, 70, 258–268.

Norquist, J. M., Fitzpatrick, R., Dawson, J. and Jenkinson, C. (2004). Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Medical Care*, 42(1 Suppl), I25–I36.

Numbers, K., *et al.* (2020). Participant and informant memory-specific cognitive complaints predict future decline and incident dementia: findings from the Sydney Memory and Ageing Study. *PLOS ONE*, 15, e0232961.

Park, M. H. (2017). Informant questionnaire on cognitive decline in the elderly (IQCODE) for classifying cognitive dysfunction as cognitively normal, mild cognitive impairment, and dementia. *International Psychogeriatrics*, 29, 1461–1467.

Perroco, T. R. *et al.* (2008). Short IQCODE as a screening tool for MCI and dementia: preliminary results. *Dementia and Neuropsychologia*, 2, 300–304.

Phung, T. K. T. *et al.* (2015). Performance of the 16-item informant questionnaire on cognitive decline for the elderly (IQCODE) in an Arabic-speaking older population. *Dementia and Geriatric Cognitive Disorders*, 40, 276–289.

Ponds, R. W. and Jolles, J. (1996). Memory complaints in elderly people: the role of memory abilities, metamemory, depression, and personality. *Educational Gerontology: An International Quarterly*, 22, 341–357.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. University of Californis Press. Symposium conducted at the meeting of the Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, California.

Sachdev, P. S. *et al.* (2010). 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. *International Psychogeriatrics*, 22, 1248–1264.

Sierra-Rio, A. *et al.* (2016). Cerebrospinal fluid biomarkers predict clinical evolution in patients with subjective cognitive decline and mild cognitive impairment. *Neurodegenerative Diseases*, 16, 69–76.

Skoog, I. *et al.* (2017). Decreasing prevalence of dementia in 85-year olds examined 22 years apart: the influence of education and stroke. *Scientific Reports*, 7, 1–8.

Slavin, M. J. *et al.* (2015). Predicting cognitive, functional, and diagnostic change over 4 years using baseline subjective cognitive complaints in the Sydney Memory and Ageing Study. *The American Journal of Geriatric Psychiatry*, 23, 906–914.

Smith, E. V. Jr (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3, 205–231.

Striepens, N. *et al.* (2010). Volume loss of the medial temporal lobe structures in subjective memory impairment. *Dementia and Geriatric Cognitive Disorders*, 29, 75–81.

Stucki, G., Daltroy, L., Katz, J., Johannesson, M. and Liang, M. H. (1996). Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts. *Journal of Clinical Epidemiology*, 49, 711–717.

Tang, W. K. *et al.* (2004). The scoring scheme of the informant questionnaire on cognitive decline in the elderly needs revision: results of rasch analysis. *Dementia and Geriatric Cognitive Disorders*, 18, 250–256.

Tennant, A. and Conaghan, P. G. (2007). 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 and Research*, 57, 1358–1362.

Truong, Q. *et al.* (2021). Applying generalizability theory to examine assessments of subjective cognitive complaints: whose reports should we rely on–participant versus informant? *International Psychogeriatrics*, 1–11. doi: 10.1017/S1041610221000363