

Applying Generalizability Theory to Examine Assessments of Subjective Cognitive Complaints: Whose Reports Should We Rely on - Participant versus Informant?

Running Title: Assessments of Subjective Cognitive Complaints

Journal:	<i>International Psychogeriatrics</i>
Manuscript ID	IPG-12-20-719.R2
Manuscript Type:	Original Research Article
Date Submitted by the Author:	23-Feb-2021
Complete List of Authors:	Truong, Quoc; University of Waikato Choo, Carol; University of Waikato Numbers, Katya; University of New South Wales, Psychiatry Merkin, Alexander; Auckland University of Technology Brodaty, Henry; UNSW, Dementia Collaborative Research Centre; Prince of Wales Hospital, Aged Care Psychiatry Kochan, Nicole; UNSW, Brain and Ageing Research Program, School of Psychiatry Sachdev, Perminder; UNSW, Brain and Ageing Research Program, School of Psychiatry Feigin, VALery; Auckland University of Technology Medvedev, Oleg; University of Waikato, Department of Psychology
Keywords:	Aging, Cognitive assessment, Dementia, Mental capacity, Research Design and Methodology
Abstract:	<p>Objectives: This study aimed to apply the Generalizability Theory (G-Theory) to investigate dynamic and enduring patterns of Subjective Cognitive Complaints (SCC), and sources of measurement errors and reliability of two widely used SCC assessment tools.</p> <p>Design: G-Theory was applied to assessment scales using longitudinal measurement design with five assessments spanning 10 years of follow-up.</p> <p>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).</p> <p>Participants: The sample included 232 participants aged 70 years and older, and 232 associated informants. Participants were predominantly white Europeans (97.8%). The sample of informants included 76 males (32.8%), 153 females (65.9%), and their age ranged from 27 to 86 years, with a mean age of 61.3 years (SD=14.38).</p> <p>Measurements: The Memory Complaint Questionnaire (MAC-Q) and the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE).</p> <p>Results: The IQCODE demonstrated strong reliability in measuring enduring patterns of SCC with $G=0.86$. Marginally acceptable reliability of the 6-item MAC-Q ($G=0.77-0.80$) was optimized by removing one item resulting in $G=0.80-0.81$. Most items of both assessments were measuring enduring SCC with exception of one dynamic MAC-Q item. The IQCODE significantly predicted global cognition scores and risk of dementia incident across all occasions, while MAC-Q scores were only significant predictors on some occasions.</p>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

	Conclusions: While both informants' (IQCODE) and self-reported (MAC-Q) SCC scores were generalizable across sample population and occasions, self-reported (MAC-Q) scores may be less accurate in predicting cognitive ability and diagnosis of each individual.



Applying Generalisability Theory to Examine Assessments of Subjective Cognitive Complaints: Whose Reports Should We Rely on - Participant versus Informant?

Running Title: Assessments of Subjective Cognitive Complaints

Abstract

Objectives: This study aimed to apply the Generalizability Theory (G-Theory) to investigate dynamic and enduring patterns of Subjective Cognitive Complaints (SCC), and sources of measurement errors and reliability of two widely used SCC assessment tools.

Design: G-Theory was applied to assessment scales using longitudinal measurement design with five assessments spanning 10 years of 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 232 participants aged 70 years and older, and 232 associated informants. Participants were predominantly white Europeans (97.8%). The sample of informants included 76 males (32.8%), 153 females (65.9%), and their age ranged from 27 to 86 years, with a mean age of 61.3 years (SD=14.38).

Measurements: The Memory Complaint Questionnaire (MAC-Q) and the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE).

Results: The IQCODE demonstrated strong reliability in measuring enduring patterns of SCC with $G=0.86$. Marginally acceptable reliability of the 6-item MAC-Q ($G=0.77-0.80$) was optimized by removing one item resulting in $G=0.80-0.81$. Most items of both assessments were measuring enduring SCC with exception of one dynamic MAC-Q item. The IQCODE significantly predicted global cognition scores and risk of dementia incident across all occasions, while MAC-Q scores were only significant predictors on some occasions.

Conclusions: While both informants' (IQCODE) and self-reported (MAC-Q) SCC scores were generalizable across sample population and occasions, self-reported (MAC-Q) scores may be less accurate in predicting cognitive ability and diagnosis of each individual.

Applying Generalizability Theory to Examine Assessments of Subjective Cognitive Complaints: Whose Reports Should We Rely on - Participant versus Informant?

Introduction

Older adults often report subjective cognitive complaints (SCC), which relate to an individual's self-experience of cognitive deterioration (Hildreth and Church, 2015). Currently, SCC contribute to the criteria for a diagnosis of mild cognitive impairment (MCI) (Petersen, 2016; Winblad et al., 2004), and may be considered as the earliest detectable stage of pre-clinical dementia (Jonker et al., 2000; Mitchell et al., 2014). SCC can be self-reported or reported by informants (e.g. family member or friend) with the advantage of capturing daily cognitive and memory changes that standardized neuropsychological tests may not detect (Brodaty et al., 2002; Jorm et al., 1991; Numbers et al., 2020).

Despite the potential benefits of SCC assessments, it remains questionable as to whether self-reported SCC reliably predict objective cognitive performance and/or dementia incident. A relationship between self-reported SCC and cognitive impairment ranges from negligible (e.g., Burmester et al., 2016; Crumley et al. 2014) to none (e.g., Lenihan et al., 2012; Reid and MacLulich, 2006). One explanation for such inconsistency may be the influence of mood and certain personality traits on complaining behaviors (Ponds and Jolles, 1996). It is well established that subjective impressions of decline are exacerbated by depression and anxiety, as well as personality traits such as neuroticism and conscientiousness (Reid and MacLulich, 2006). Therefore, subjective reports of cognitive ability provided by close informants may present more reliable approximation of objective cognitive performance (Slavin et al., 2015), and future cognitive decline (Caselli et al., 2014). Furthermore, in the clinical setting, informant-reported SCC are often increasingly relied upon as individuals progress through pre-clinical stages of dementia, and begin losing insight into their cognitive changes over the debilitating course of dementia (American Psychiatric Association, 1994). However, no empirical examination was conducted to

1
2
3 date using an appropriate methodology to investigate whose reports (i.e. participants' or
4 informants') are more reliable, and at what stage researchers and clinicians should rely on which
5 reports. As dementia is typically marked by insidious onset and gradual progression, a longitudinal
6 design will be ideal in tracking any cognitive change over time (American Psychiatric Association,
7 1994).
8
9
10
11
12

13
14
15
16 Moreover, it is important to differentiate reliably between dynamic and enduring SCC
17 patterns over longer time. A reliable trait measure would reflect enduring changes overtime (e.g.
18 alterations in long-term subjective cognition) and remain unaffected by individual's transient
19 changes (e.g. mood or current stress level). Conversely, a state measure would be sensitive to
20 dynamic changes, which may confound assessment of long-term subjective cognition. While the
21 widely used SCC measures such as the Memory Complaint Questionnaire (MAC-Q) (Crook et
22 al., 1992) and the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) (Jorm,
23 1994) have good internal consistency, it does not support their ability to distinguish between
24 enduring and dynamic patterns of SCC. Internal consistency coefficients (e.g. Cronbach's alpha)
25 are not appropriate to estimate temporal reliability of scales because they only estimate
26 consistency or inter-correlations between individual items at one time point. Moreover, test-retest
27 reliability coefficients often used to distinguish between dynamic and enduring patterns have
28 limited accuracy because these are merely correlations between total scale scores at two different
29 times (e.g. Time 1 and Time 2). For example, if a person improves on one symptom but gets
30 worse on another, the total score remains the same without reflecting clinically important changes.
31 Therefore, these coefficients do not account for variability of individual items over time and other
32 sources of measurement error such as the effects of item, occasion, person, and their interactions
33 (Bloch and Norman, 2012; Medvedev et al., 2020). For instance, a response to an item may
34 depend on assessment occasion rather than changes in individual's performance. A
35 comprehensive estimation of reliability is therefore required and Generalizability Theory (G-
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

1
2
3 Theory) was advocated as the most appropriate method to investigate dynamic and enduring
4 patterns in a measure, and examine reliability and generalizability of assessment scores
5 (Medvedev et al., 2017; Truong et al., 2020).
6
7
8
9

10
11 G-Theory is a successor of Classical Test Theory (CTT), and is particularly well suited to
12 examine the overall reliability of psychometric instruments (Shavelson and Webb, 1991; Brennan,
13 2010). While CTT postulates that any measurement consists of true variance and error variance
14 presented as a single factor, G-Theory utilizes ANOVA to estimate all possible sources of error
15 variance that may affect the main outcome variable, as well as the accuracy of the measurement
16 itself (Allen and Yen, 2001; Cronbach et al., 1963). Furthermore, CTT evaluates the reliability of
17 a measure at only one aspect (e.g. internal consistency) at a time, or examines the distinction
18 between dynamic and enduring patterns of a measure using test-retest reliability coefficients. G-
19 Theory extends CTT and simultaneously examines all potential sources of error variance that may
20 influence reliability such as person, scale items, occasion, and all their interactions (Medvedev et
21 al., 2017; Shavelson et al., 1989). Many studies have demonstrated applicability of G-theory as
22 the most appropriate method for estimating the overall reliability and generalizability of
23 assessment scores and distinction between dynamic and enduring patterns in a measure
24 (Arterberry et al., 2014; Medvedev et al., 2017; Paterson et al., 2018; Truong et al., 2020).
25 Therefore, applying G-Theory can be useful to examine and improve the precision of a
26 psychometric instrument, as well as differentiate between enduring and dynamic patterns
27 reflected by such measure.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

47 The aim of the current study was to apply G-Theory to examine reliability and distinguish
48 between dynamic and enduring patterns in the self-report MAC-Q and informants IQCODE SCC
49 assessment tools. A longitudinal design was utilized with participants assessed at five occasions,
50 separated by 2-4 years intervals. Application of G-Theory involved two parts: a Generalizability
51 study (G-study) and a Decision study (D-study). The purpose of the G-study was to examine the
52
53
54
55
56
57

1
2
3 overall generalizability of the MAC-Q and IQCODE and evaluate sources of error variance in each
4
5 measure. D-study aimed to subsequently evaluate psychometric properties of individual items of
6
7 these two scales, and to manipulate measurement design to optimize the reliability of
8
9 measurement (Shavelson et al., 1989; Cardinet et al., 2011). We also aimed to evaluate the utility
10
11 of these scales for predicting incident dementia and global cognition scores.
12
13

14 15 16 **Method**

17 18 19 20 ***Participants***

21
22
23
24 Community-dwelling older adults aged 70 – 90 years, living in the Eastern Suburbs of
25
26 Sydney, Australia, were selected via the electoral roll and invited to participate in the Sydney
27
28 Memory and Ageing Study (MAS) (Sachdev et al., 2010). Of 8,914 individuals invited to participate,
29
30 1,037 participants were included in the baseline sample (occasion 1). Inclusion criteria were the
31
32 ability to speak and write English sufficiently well to complete a psychometric assessment and
33
34 self-report questionnaires. Exclusion criteria were any major psychiatric diagnoses, acute
35
36 psychotic symptoms, current diagnosis of multiple sclerosis, motor neuron disease,
37
38 developmental disability, progressive malignancy, and/or dementia. All participants provided
39
40 written consent to participate in this study, which was approved by the University of New South
41
42 Wales Human Ethics Review Committee (HC 05037, 09382, 14327). More detailed methods of
43
44 recruitment and baseline demographics have been previously described by Sachdev and
45
46 colleagues (Sachdev et al., 2010). Of the 1037 participants included in the present study, 1,009
47
48 (97.3%) had an informant. Informants were selected by nominations of participants. Informants
49
50 answered questions relating to the participant's memory, thinking, and daily functioning. Qualified
51
52 informants were those who had at least 1 hour of contact with the participant per week; on average
53
54 they had 8.3 hours of weekly contact. All participants and informants provided written consent to
55
56
57

1
2
3 participate in this study, which was approved by the University of New South Wales Human Ethics
4
5 Review Committee (HC 05037, 09382, 14327).
6
7
8

9
10 Figure1 presents consort diagram including the number of MAC-Q and IQCODE reports,
11 and the number of participants diagnosed with dementia along with computed their global
12 cognition scores, for each occasion (wave). Of the 1,009 participants with informants, 232 (23%)
13 had reports of MAC-Q and IQCODE at all five occasions and were included in the G-analyses.
14 We excluded participants' with informants (77%) whose MAC-Q and IQCODE were incomplete
15 at one or more occasions. The MAC-Q or IQCODE data were missing at some waves because
16 either the participant or informant was not contactable or was not able to do the assessment at
17 that wave. In some instances, participants were too ill or advanced in dementia to answer
18 questions in later waves (Informant Only), and in others, participants simply did not have an
19 informant who was willing to complete an interview or questionnaire on their behalf. The ethnicity
20 of the extracted sample was predominantly white Europeans (97.8%); the remaining sample was
21 0.4% other, and 1.7 % unrevealed. Informants from the extracted sample were 76 males (32.8%),
22 153 females (65.9%), and their age ranged from 27 to 86 years, with a mean age of 61.3 years
23 (SD=14.38). Missing responses per item of either the MAC-Q or the IQCODE in the extracted
24 sample comprised less than 0.05% which were negligible and thus substituted by mean
25 imputation at each respective wave (Huisman, 2000). This sample size of 232 participants
26 exceeded the required sample size of 84 participants for repeated measures ANOVA over five
27 occasions needed to accomplish the power (1- β) of 0.95 to detect effect size of 0.15 under p value
28 of 0.05.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 <Insert Figure 1 here>
53
54
55

56 **Measures**

57
58
59
60

1
2
3
4
5 The MAC-Q (Crook et al., 1992) is a well-validated unidimensional 6-item questionnaire.
6
7 Internal consistency Cronbach's alpha of the MAC-Q was reported in the range from 0.57 to 0.88
8
9 with most studies indicating acceptable values and confirming unidimensionality of the scale
10
11 (Buckley et al., 2013, Crook et al., 1992, Reid et al., 2012). The MAC-Q asked participants to rate
12
13 themselves compared to how they previously performed on several everyday memory tasks (e.g.
14
15 difficulty remembering names; see Supplemental Materials for a full list of items). At occasion 1,
16
17 participants received the conventional MAC-Q wording "How would you rate yourself compared
18
19 to 5 years ago", but for each subsequent occasion, the wording was changed to "How would you
20
21 rate yourself compared to 2 years ago" to capture the intervening time between assessments.
22
23 Participants rated themselves for each item on a scale of 1 to 5; total score range from 5 to 30,
24
25 with higher scores indicating greater subjective memory loss.
26
27
28
29

30
31 The IQCODE (Jorm, 1994) consists of 16 items that asks informants to report on their
32
33 perceived changes of the participant's cognition and functioning. Each item is scored on a 5-point
34
35 Likert scale with options ranging from 1 = "much improved" to 5 = "much worse". The IQCODE is
36
37 completed by informants who are well-known to the individual (Harrison et al., 2016) and has
38
39 been shown to reliably predict incident dementia (Numbers et al., 2020). The original IQCODE
40
41 consists of 26 questions/items. An abbreviated version of the IQCODE consisting of 16 items has
42
43 been found to perform as reliably as the original version (Jorm, 1994, Jorm, 2004), with a number
44
45 of studies confirming high internal consistency (Cronbach's alpha = 0.93 to 0.97) (Harrison et al.,
46
47 2015, Phung et al., 2015, Tang et al., 2004) and a superior ability to predict incident dementia
48
49 (Perroco et al., 2008; Park, 2017).
50
51
52
53

54 **Dementia Diagnosis**

55
56
57
58
59
60

1
2
3 Clinical diagnoses were performed for all occasions (10-year follow-up). At occasion 1,
4 and at each two-year follow-up, individuals were brought to a consensus review meeting where
5 at least three clinicians from a panel of neuropsychiatrists, psychogeriatricians, and
6 neuropsychologists discussed all available clinical, neuropsychological, laboratory and imaging
7 data to reach a consensus diagnosis. A diagnosis of dementia was based on the Diagnostic and
8 Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (American Psychology
9 Association, 1994) - that is, the development of one or more cognitive deficit(s) that represent a
10 decline from a previous level of performance and were sufficiently severe as to cause impairment
11 in functioning (Bayer IADL scale score ≥ 3.0). Individuals who did not receive a dementia
12 diagnosis were classified as "not dementia" at each occasion, and no dementia cases were
13 present at occasion 1 as this was an exclusionary criterion (Sachdev et al., 2010).
14
15
16
17
18
19
20
21
22
23
24
25
26
27

28 **Objective Cognitive Performance**

29
30
31
32 Comprehensive cognitive data were available for occasion 1 to occasion 4 (6-year follow
33 up), only. Cognitive performance over these first four occasions was assessed using a
34 comprehensive neuropsychological test battery which comprised 10 tests that measured the
35 domains of attention/processing speed, language, executive function, visuospatial ability, and
36 memory (see Supplementary Table S1). Domain and global cognition composites were computed
37 as standardized z-scores as follows. Firstly, raw test scores were converted to z-scores using the
38 baseline means and standard deviations (SDs) of a reference group which comprised 732 MAS
39 participants classified as cognitively healthy at occasion 1 (native English speakers with a Mini-
40 Mental State Examination score of 24 or above, no evidence of dementia or current depression,
41 no history of delusions or hallucinations, and no major neurological disease, significant head
42 injuries, progressive malignancies or CNS medications). Of the 732 participants (ages ranged
43 from 70.29 to 90.80 years with $M=78.57$, $SD=4.72$), 219 (29.9%) completed tertiary qualification,
44
45
46
47
48
49
50
51
52
53
54
55
56
57

1
2
3 128 (17.5%) completed high school and/or diploma, 350 (47.8%) were not completed high
4 school/diploma, 23 (3.2%) incomplete tertiary qualification, and 12 (1.6%) completed primary
5 school. Secondly, composite domain scores were formed by averaging the z-scores of the
6 component tests (as defined above), apart from the visuospatial domain which was represented
7 by a single test. Each domain composite was standardized by transforming so that the mean and
8 standard deviation of the baseline cognitively healthy group were 0 and 1, respectively. Finally,
9 global cognition scores for each occasion were calculated by averaging the domain z-scores, and
10 again transforming these scores so that the means and standard deviations for the baseline
11 reference group were 0 and 1, respectively.
12
13
14
15
16
17
18
19
20
21
22
23

24 **Data analyses**

25
26
27
28 EduG 6.1-e software (Swiss Society for Research in Education Working Group, 2006) was
29 used to conduct Generalizability analyses by following the guidelines described in Truong et al.
30 (2020). Both G-study and D-study used two-facet design (person by item by occasion) which item
31 (I) and occasion (O) were two facets of interest (instrumentation facets) and person (P) was the
32 object of measurement (differentiation facet), expressed as $P \times I \times O$ (Cardinet et al., 2011, Peng,
33 2007, Truong et al., 2020, Vispoel et al., 2018). The facet of I was fixed because the same items
34 of assessments were used across all participants and all occasions, whereas the P and O facets
35 were infinite. Besides that, the facet P was not a source of error and in a study employing G-
36 Theory method, all error variances are counted as 100% after controlling for person variance (P),
37 which reflects true differences between persons (Cardinet et al., 2011). G-Theory estimates for
38 the design of person by item by occasion, express as $P \times I \times O$ were calculated using formulae
39 included in Supplementary Table S2 (Shavelson et al., 1989).
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

There are two reliability coefficients, relative G-coefficient (G_r) and absolute G-coefficient (G_a), for the object of measurement (person) in a generalizability study. The relative model of measurement is based on a norm-referenced manner in which a person's assessment score is compared against the scores of others (Peng, 2007; Vispoel et al., 2018). G_r accounts for a relative error variance (σ_{δ}^2) which is related to the I facet (object of measurement) that may affect a relative measurement (e.g., interaction between person and occasion- PxQ, and interaction between person and item- Pxl) and includes divisions by desired sample sizes (Shavelson and Webb, 1991). Both G coefficients are estimating reliability of an enduring pattern of a measurement if the person (P) is differentiation facet. Specifically, G_r of 0.80 or higher is determined as good reliability of assessment score (Cardinet et al., 2011) while G_a above 0.70 is considered as acceptable reliability (Truong et al., 2020; Arterberry et al., 2014). Both a state component index (SCI) and trait component index (TCI) were obtained, which represent the variance proportion attributed to a dynamic (state) and an enduring (trait) pattern in a measure (Medvedev et al., 2017). SCI of 0.60 or higher (TCI<0.40) would indicate that variance is reflecting a dynamic pattern. On the contrary, TCI above 0.60 (SCI<0.40) would signify a variance is reflecting an enduring pattern. In the D-study, variance components were computed for each individual item and effects of removing facets levels were examined to optimize the reliability of the MAC-Q and the IQCODE.

IBM SPSS Statistics 25 software was used to compute estimates that related to CTT approaches and descriptive statistics including means, standard deviation (SD), Cronbach's alpha, and intraclass correlation coefficient (ICC) for the IQCODE and the MAC-Q. Logistic regression analyses were conducted to examine how the IQCODE and the MAC-Q were able to predict the incidence of dementia across occasions 2-5. Three logistic regression models were carried out for each occasion. Each logistic regression model involved the outcome variable of

dementia diagnose with one predictor of either MAC-Q scores or IQCODE scores at the same occasion. Moreover, linear regression analyses were also used to estimate ability of these measures to predict global cognition scores across the first four occasions. Three independent linear regression models were conducted at each occasion with the outcome variable of global cognition scores and either MAC-Q scores or IQCODE scores as a predictor. Prior to all regression analyses, assumption tests were conducted to screen for potential violations.

Results

G-Study

The variance components attributed to person (P), item (I), and occasion (O), and their interactions (Pxl, PxO, lxO, PxlxO) together with generalizability coefficients and state and trait component indices are presented in Table 1 for the MAC-Q and Table 2 for the IQCODE. The IQCODE showed better reliability and generalizability of scores across persons and occasions, with both relative and absolute G coefficients of 0.86, 95%CI [0.84; 0.88]. Measurement error was predominantly explained by Pxl and PxO interactions for the IQCODE, which together explained 77.1% of the total error variance after accounting for the true person variance. Slightly lower, but still acceptable values ($G_r=0.80$, 95%CI [0.77; 0.83]; $G_a=0.77$, 95%CI [0.73; 0.81]) were observed for the 6-item MAC-Q, with the main source of error variance due to the Pxl interaction explaining 35.7% of the total error variance. Consistent with reliability estimates, TCI values were 0.96 for the MAC-Q and 0.95 for the IQCODE indicating that both instruments reliably assess enduring patterns of SCC.

<Insert Table 1 and 2 here>

D-Study

A series of generalizability analysis were conducted to obtain variance components for each individual items of the MAC-Q and IQCODE. The item-level estimates for variance of person, occasion, and person-occasion interaction, together with computed SCI, are included in Supplementary Table S3. There was only one MAC-Q item reflecting high sensitivity for transitory changes in SCC patterns over time; “item e: Remembering the item(s) you intended to buy when you arrive at the supermarket store or pharmacy?”, which had the highest SCI of 0.66. The other five items of the MAC-Q revealed an SCI range from 0.15 to 0.45 indicating a lower proportion of variance associated with dynamic changes in SCC over time. However, all IQCODE items reflected predominantly enduring patterns of SCC.

Six additional generalizability analyses were conducted by excluding one item at a time for the MAC-Q, as we expected that this may result in improving the reliability of the scale in measuring enduring patterns of SCC (Table 1). The first analysis involved removing the first item (item a), with subsequent analyses removing one item at a time and examining reliability. Removing the final item (f) of the MAC-Q “In general, how would you describe your memory as compared to 10 years ago?” was the only analysis that resulted in improvement of both relative and absolute G coefficients above 0.80 bench mark suggesting that the 5-item MAC-Q (i.e. MAC-Q without item f) has better reliability compared to the 6-item MAC-Q. Next analyses involved removing one occasion at a time for the MAC-Q (Supplementary Table S4) to examine how this affects the reliability of the scale. Removing any occasion only slightly decreased both G coefficients, which remained in the acceptable range.

Additional G-analyses were conducted on the IQCODE (Table 2), which involved removing items more sensitive to dynamic SCC with $SCI \geq 0.40$ in attempt to optimize reliability.

1
2
3 This resulted in lower G-coefficients compared to the original IQCODE. Additionally, removing
4 one occasion at a time for the IQCODE only slightly decreased G-coefficients. These findings
5 together support reliability of the IQCODE with the current measurement design.
6
7
8
9

10 11 ***Classical Test Theory Analyses*** 12

13
14
15
16 Descriptive statistics for the 5-item MAC-Q, the 6-item MAC-Q, and the IQCODE at five occasions
17 are presented in Supplementary Table S5. The internal consistency Cronbach's alpha of the 6-
18 item MAC-Q was fair to good over five occasions and ranged between 0.57 and 0.76, which was
19 consistent with values reported by other studies (Buckley et al., 2013, Crook et al., 1992, Reid et
20 al., 2012). Given lower Cronbach alpha, we explored a possibility of multidimensionality that could
21 impact on internal consistency of the MAC-Q with the current dataset using full sample at occasion
22 1 (n=1011) by applying exploratory factor analysis. The results showed only one factor with
23 eigenvalue >1 using Kaiser Criteria and the scree plot clearly indicated elbow after one factor with
24 loading on the first principle component ranging from 0.53 to 0.75 supporting unidimensionality of
25 the MAC-Q. Temporal stability was supported by intraclass correlation coefficient (ICC) of 0.84
26 across all occasions. The mean scores of the 5-item MAC-Q were only significantly different
27 between occasion 1 and 2, while that of the 6-item MAC-Q were significantly different between
28 occasion 1 and occasions 2 and 3. The IQCODE demonstrated higher internal consistency with
29 Cronbach's alphas ranging from 0.84 to 0.95, though the ICC of 0.70 was lower than both MAC-
30 Q scales. Overall, the MAC-Q and IQCODE scales showed acceptable to high internal reliability
31 and acceptable temporal reliability for a measure of enduring patterns over time, which is
32 consistent with G-study results.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 To evaluate predictive validity of the original MAC-Q (6-item), the shortened MAC-Q (5-
4 item) and the IQCODE, a series of binary logistic regression analyses were conducted to predict
5 risk of incident dementia. Table 3 presents coefficients for the models' predictors, as well as model
6 percentage correct (MPC), for the three SCC measures across the 4 follow-up occasions
7 (occasion 1 was excluded because all participants were initially healthy). Prior to the analyses,
8 assumption testing was conducted for all the models and did not indicate any violations. The
9 Hosmer and Lemeshow tests indicated good-fit for these logistic regression models (all
10 p 's > 0.05). Accuracy of all models across occasions were ranging from 78.2% to 98.2% in their
11 predictions of incidence of dementia. The IQCODE significantly predicted incident dementia at all
12 examined occasions, with all p 's < 0.001. Whereas, the two versions of the MAC-Q (i.e., 5-item
13 vs. 6-item) were only significant predictors of dementia incident on occasions 3 and 4 with p 's ≤
14 0.02 but not at occasions 2 and 5 (p 's ≥ 0.30).
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

31 <Insert Table 3 here>
32
33
34

35 Table 4 presents a series of linear regression analyses conducted to determine the
36 relationship between predictors such as the two MAC-Q scales and the IQCODE, and the
37 outcome measured as participants' global cognition scores at the first four occasions. Several
38 assumptions were evaluated for these linear regression models prior to the interpretation of the
39 results. The data was distributed close to normal with skewness values for all variables in the
40 models ranging from -1.03 (dementia diagnose at occasion 4) to 2.12 (5-item MAC-Q scores at
41 occasion 5). Inspection of normal probability plots of regression standardized residuals also
42 indicated that these variables were normally distributed. The scatterplots of standardized
43 residuals were compared against standardized predicted values and also revealed that these
44 variables met the assumptions of linearity, homoscedasticity of residuals, and were free from
45 univariate outliers. Finally, there were no multicollinearity issue because only one predictor was
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 independently tested for each model. The IQCODE significantly predicted global cognition scores
4 for all four occasions, with all p 's < 0.01 while both MAC-Q versions were not significantly
5 associated with global cognition at the first occasion (p 's > 0.05)."
6
7
8
9

10
11 <Insert Table 4 here>
12
13
14

15 Discussion

16
17 The aim of the current study was to evaluate and optimize reliability, and distinguish
18 between dynamic and enduring patterns in the MAC-Q (self-report) and IQCODE (informants)
19 SCC measures using G-Theory. The results showed that the optimized 5-item MAC-Q and the
20 IQCODE were reliable in measuring enduring pattern of SCC with G-coefficients of 0.80 and
21 higher, and index of trait (TCI) above 0.94, suggesting that their scores are generalizable across
22 sample population and occasions. In line with previous research (Slavin et al., 2015), we found
23 that the IQCODE SCC scores significantly predicted risk of dementia incident and global cognition
24 across all occasions while the MAC-Q scores were only significant predictors on some occasions.
25 However, these results should be interpreted with caution due to differences in length and format
26 between the IQCODE and the MAC-Q. Together our findings suggest that the MAC-Q reliably
27 measures individual levels of SCC, but these self-reported SCC may be less accurate in reflecting
28 cognitive abilities and diagnosis of each individual. It is possible that the MAC-Q tend to reflect
29 individual tendencies to report complaints (e.g., about their self-perceived memory errors) rather
30 than their actual cognitive capacities. In other words, some people may have stronger tendency
31 to ruminate on everyday memory errors or lean towards complaining behavior, which may not
32 necessary reflect their actual cognitive abilities. Consistent with the recent clinical literature, SCC
33 may be related to anxiety and stress in individuals with normal cognition (Chin et al., 2019). The
34 outcome of our study has clinical implications, which underscore the importance for clinicians to
35 seek corroboratory evidence from knowledgeable informants in their follow-up of aging patients.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

1
2
3 This in turn, could help in detection of Mild Cognitive Impairment, which is the pre-clinical stage
4 of the trajectory of cognitive decline, and would assist in ongoing clinical management and
5 planning, as once dementia is diagnosed, it runs a debilitating course (Langa and Levine, 2014).
6
7
8

9 A D-study was conducted to examine psychometric properties of individual items of the
10 MAC-Q and IQCODE in an effort to optimize reliability of the measurement. Results showed that
11 most individual items of the IQCODE and MAC-Q measured enduring patterns of SCC, except
12 item-e of the MAC-Q (item – e: “Remembering the item(s) you intended to buy when you arrive
13 at the supermarket store or pharmacy”). However, removing this item did not improve the reliability
14 of the MAC-Q in measuring enduring pattern of SCC. Similar results were found when removing
15 each item of the MAC-Q one at a time, with the exception of the last item (item-f: “In general, how
16 would you describe your memory as compared to 10 years ago?”). Removing this last item
17 boosted the marginally acceptable reliability of the 6-item MAC-Q ($G=0.77-0.80$) up to $G=0.80-$
18 0.81 in the optimized 5-item MAC-Q. No reliability improvements were achieved by manipulating
19 measurement design of the IQCODE suggesting optimal reliability of the scale in the current
20 measurement design.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

37 **Strengths and limitations**

38
39 The main strength of the study was to apply the comprehensive methodology of G-Theory
40 to a relatively large sample in a longitudinal study spanning over 10 years. However, limitations
41 need to be acknowledged. The study was conducted with participants recruited from a relatively
42 small catchment area in Sydney, Australia. Moreover, the participants belonged to a
43 predominantly White (European) ethnic group, and the generalizability to other ethnicities is
44 questionable. Recent research suggests that cultural variations contribute to vulnerabilities and
45 resilience across a range of health issues (Choo et al., 2017). As such, it would be beneficial to
46 replicate these analyses on samples comprising other ethnicities, including culturally and
47 linguistically diverse groups. This study aimed to analyze data from 5 occasions with equal
48
49
50
51
52
53
54
55
56
57

1
2
3 intervals of 2 years; however, the interval between occasion 4 and occasion 5 was 4 years, as
4
5 MAC-Q data was missing for wave 5 (8-year follow-up) assessments. Future studies should
6
7 endeavor to replicate these analyses using equal intervals between occasions.
8

9
10 The findings of this study added evidence supporting the benefits of using the informant
11
12 SCC report. However, due to differences between the IQCODE and the MAC-Q format, more
13
14 accurate comparison between informants' reports measured by the IQCODE and self-reports
15
16 could be achieved using the self-report version of the IQCODE - the Informant Questionnaire on
17
18 Cognitive Decline- Self-report (IQCODE-SR) (Jansen et al., 2008). Therefore, further studies are
19
20 warranted to compare the IQCODE with the IQCODE-SR.
21

22
23 In addition, this study did not control for demographic variables (e.g., age, sex,
24
25 socioeconomic and education status), mood or personality in both sets of regression analyses.
26
27 Nevertheless, the results of G-analyses indicated that the IQCODE and both versions of MAC-Q
28
29 were measuring enduring pattern of SCC and were less affected by dynamic and transient
30
31 conditions such as mood. Notably, less than 20% of variance was explained by error due to
32
33 temporal factor and interactions ($G=0.80-0.86$).
34

35 **Conclusion**

36
37 The findings of this study indicated that the IQCODE and MAC-Q assessment scores were
38
39 generalizable across sample population and occasions and captured enduring patterns of SCC
40
41 over 10-years. The optimized 5-item MAC-Q was superior to the original 6-item scale when
42
43 assessing SCC over time. While clinicians and researchers could rely on both participants and
44
45 informants' SCC reports of the IQCODE and the MAC-Q, self-reported (MAC-Q) scores may be
46
47 less accurate in predicting cognitive ability and diagnosis of each individual.
48
49

50 **Compliance with Ethical Standards**

51
52 The study complied with the guidelines of the university ethics committee, which are
53
54 consistent with internationally accepted ethical standards.
55
56

Conflict of Interest

None

Description of Author Roles

Q. Truong designed the study, conducted statistical analyses and wrote the manuscript. C. Choo supervised the study and edited the manuscript. K. Numbers collaborated with data collection and study design, and edited the manuscript. A. Merkin and V. Feigin collaborated with writing and editing the manuscript. N. Kochan assisted with designing instruments, collecting data and editing the manuscript. P. Sachdev and H. Brodaty sourced funding, supervised data collection, collaborated with writing and editing the manuscript. O. Medvedev supervised the study and data analyses, and edited the manuscript.

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>

References

- Allen, M. J. and Yen, W. M. (2001). Introduction to measurement theory: Waveland Press.
- American Psychology Association (1994). Diagnostic and statistical manual of mental disorders. Washington.

- 1
2
3 Arterberry, B. J., Martens, M. P., Cadigan, J. M. and Rohrer, D. (2014). Application of
4 generalizability theory to the big five inventory. *Personality and individual differences*, 69, 98-
5 103.
6
7
8
9
10 Bloch, R. & Norman, G. (2012). Generalizability theory for the perplexed: a practical introduction
11 and guide: AMEE Guide No. 68. *Medical teacher*, 34(11), 960-992.
12
13
14 Brennan, R. L. (2010). Generalizability theory and classical test theory. *Applied Measurement in*
15 *Education*, 24(1), 1-21.
16
17
18
19 Brodaty, H., *et al.* (2002). The GPCOG: a new screening test for dementia designed for general
20 practice. *Journal of the American Geriatrics Society*, 50(3), 530-534.
21
22
23
24 Buckley, R., *et al.* (2013). Factors affecting subjective memory complaints in the AIBL aging
25 study: biomarkers, memory, affect, and age. *International Psychogeriatrics*, 25 (8) 1307-1315.
26
27
28
29 Burmester, B., Leathem, J. and Merrick, P. (2016). Subjective Cognitive Complaints and
30 Objective Cognitive Function in Aging: A Systematic Review and Meta-Analysis of Recent
31 Cross-Sectional Findings. *Neuropsychol Rev*, 26(4), 376–93.
32
33
34
35 Cardinet, J., Johnson, S. and Pini, G. (2011). Applying generalizability theory using EduG:
36 Taylor & Francis.
37
38
39
40 Caselli, R. J., *et al.* (2014). Subjective cognitive decline: self and informant comparisons.
41 *Alzheimer's & Dementia*, 10(1), 93-98.
42
43
44
45 Chin, Y., Choo, C. C. and Doshi, K. (2019). A Case of Subjective Cognitive Complaints in Older
46 Adults: Anxiety, Stress, and Aging in an Elderly Client. In *Clinical Psychology Casebook Across*
47 *the Lifespan* (pp. 71-77): Springer.
48
49
50
51
52 Choo, C. C., Harris, K. M., Chew, P. K. and Ho, R. C. (2017). Does ethnicity matter in risk and
53 protective factors for suicide attempts and suicide lethality? *PLoS One*, 12(4). doi: e0175752.
54
55
56
57
58
59
60

- 1
2
3 Cronbach, L. J., Rajaratnam, N. and Gleser, G. C. (1963). Theory of generalizability: A
4 liberalization of reliability theory. *British Journal of Statistical Psychology*, 16(2), 137-163.
5
6
7
8 Crook, T. H., Feher, E. P. and Larrabee, G. J. J. I. P. (1992). Assessment of memory complaint
9 in age-associated memory impairment: the MAC-Q. 4(2), 165-176.
10
11
12 de Jonghe, J. F. (1997). Differentiating between demented and psychiatric patients with the
13 Dutch version of the IQCODE. *International Journal of Geriatric Psychiatry*, 12(4), 462-465.
14
15
16
17 Harrison, J. K., Fearon, P., Noel-Storr, A. H., McShane, R., Stott, D. J. and Quinn, T. J. (2015).
18 Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of
19 dementia within a secondary care setting. *Cochrane Database of Systematic Reviews* (3).
20
21
22
23
24 Harrison, J. K., Stott, D. J., McShane, R., Noel-Storr, A. H., Swann-Price, R. S. and Quinn, T. J.
25 (2016). Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the early
26 diagnosis of dementia across a variety of healthcare settings. *Cochrane Database of Systematic*
27 *Reviews* (11).
28
29
30
31
32
33 Hildreth, K. L. and Church, S. J. M. C. (2015). Evaluation and management of the elderly patient
34 presenting with cognitive complaints. 99(2), 311-335.
35
36
37
38 Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Quality*
39 *and Quantity*, 34(4), 331-351.
40
41
42
43 Jansen, A.P.D., et al. (2008). Self-reports on the IQCODE in Older Adults: A Psychometric
44 Evaluation. *J Geriatr Psychiatry Neurol*, 21(2):83–92.
45
46
47 Jonker, C., Geerlings, M. I. and Schmand, B. (2000). Are memory complaints predictive for
48 dementia? A review of clinical and population-based studies. *International journal of geriatric*
49 *psychiatry*, 15(11), 983-991.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Jorm, A. (1994). A short form of the Informant Questionnaire on Cognitive Decline in the Elderly
4 (IQCODE): development and cross-validation. *Psychological medicine*, 24(1), 145-153.
5
6

7
8 Jorm, A. F. (2004). The Informant Questionnaire on cognitive decline in the elderly (IQCODE): a
9 review. *International Psychogeriatrics*, 16(3), 275.
10
11

12 Jorm, A., Scott, R., Cullen, J. and MacKinnon, A. J. P. M. (1991). Performance of the Informant
13 Questionnaire on Cognitive Decline in the Elderly (IQCODE) as a screening test for dementia.
14
15
16
17 21(3), 785-790.
18

19 Langa, K. M. and Levine, D. A. (2014). The diagnosis and management of mild cognitive
20 impairment: a clinical review. *Jama*, 312(23), 2551-2561.
21
22
23

24 Lenehan, M. E., Klekociuk, S. Z. and Summers, M. J. (2012). Absence of a relationship between
25 subjective memory complaint and objective memory impairment in mild cognitive impairment
26 (MCI): is it time to abandon subjective memory complaint as an MCI diagnostic criterion?
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

33 Medvedev, O. N., *et al.* (2020). A novel way to quantify schizophrenia symptoms in clinical trials.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

38 Medvedev, O. N., Krägeloh, C. U., Narayanan, A. and Siegert, R. J. (2017). Measuring
39 mindfulness: applying generalizability theory to distinguish between state and trait. *Mindfulness*,
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

44 Mitchell, A.J., Beaumont, H., Ferguson, D., Yadegarfar, M. and Stubbs, B. (2014). Risk of
45 dementia and mild cognitive impairment in older people with subjective memory complaints:
46 meta-analysis. *Acta Psychiatr Scand*, 130(6), 439–51.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

51 Morales, J. M., Bermejo, F., Romero, M. and Del-Ser, T. (1997). Screening of dementia in
52 community-dwelling elderly through informant report. *International Journal of Geriatric*
53
54
55
56
57
58
59
60

1
2
3 Numbers, K., Crawford, J. D., Kochan, N. A., Draper, B., Sachdev, P. S. and Brodaty, H. (2020).
4
5 Participant and informant memory-specific cognitive complaints predict future decline and
6
7 incident dementia: Findings from the Sydney Memory and Ageing Study. *PLoS One*, 15(5). doi:
8
9 e0232961.

10
11
12 Park, M. H. (2017). Informant questionnaire on cognitive decline in the elderly (IQCODE) for
13
14 classifying cognitive dysfunction as cognitively normal, mild cognitive impairment, and dementia.
15
16 *Int Psychogeriatr*, 29(9), 1461-1467.

17
18
19 Paterson, J., *et al.* (2018). Distinguishing transient versus enduring aspects of depression in
20
21 New Zealand Pacific Island children using Generalizability Theory. *Journal of affective*
22
23 *disorders*, 227, 698-704.

24
25
26 Peng, S. K. L. P. Z. (2007). Classical versus Generalizability theory of measurement.
27
28 *Examinations Research*(4), 9.

29
30
31 Perroco, T. R., *et al.* (2008). Short IQCODE as a screening tool for MCI and dementia:
32
33 preliminary results. *Dementia & Neuropsychologia*, 2(4), 300-304.

34
35
36 Petersen, R. C. (2016). Mild cognitive impairment. *CONTINUUM: Lifelong Learning in*
37
38 *Neurology*, 22 (2 Dementia), 404.

39
40
41 Phung, T. K. T., *et al.* (2015). Performance of the 16-item informant questionnaire on cognitive
42
43 decline for the elderly (IQCODE) in an Arabic-speaking older population. *Dementia and geriatric*
44
45 *cognitive disorders*, 40(5-6), 276-289.

46
47
48 Ponds, R. W. and Jolles, J. (1996). Memory complaints in elderly people: The role of memory
49
50 abilities, metamemory, depression, and personality. *Educational Gerontology: An International*
51
52 *Quarterly*, 22(4), 341-357.

53
54
55 Reid, L.M. and MacLulich, A.M.J. (2006) Subjective Memory Complaints and Cognitive
56
57 Impairment in Older People. *Dement Geriatr Cogn Disord*, 22(5-6):471-85.

1
2
3 Sachdev, P. S., *et al.* (2010). The Sydney Memory and Ageing Study (MAS): methodology and
4 baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-
5 demented cohort of Australians aged 70-90 years. *22*(8), 1248.
6
7

8
9
10 Shavelson, R. J. and Webb, N. M. (1991). *Generalizability theory: A primer* (Vol. 1): Sage.

11
12 Shavelson, R. J., Webb, N. M. and Rowley, G. L. (1989). *Generalizability theory*. *American*
13 *Psychologist*, *44*(6), 922.
14
15

16
17 Slavin, M. J., *et al.* (2015). Predicting cognitive, functional, and diagnostic change over 4 years
18 using baseline subjective cognitive complaints in the Sydney Memory and Ageing Study. *The*
19 *American Journal of Geriatric Psychiatry*, *23*(9), 906-914.
20
21
22

23
24 Swiss Society for Research in Education Working Group, (2006). *EDUG User Guide*. IRDP,
25 Euchatel, Switzerland.
26
27

28
29 Tang, W. K., *et al.* (2004). The scoring scheme of the informant questionnaire on cognitive
30 decline in the elderly needs revision: results of Rasch analysis. *Dementia and geriatric cognitive*
31 *disorders*, *18*(3-4), 250-256.
32
33
34

35
36 Truong, Q. C., Krageloh, C. U., Siegert, R. J., Landon, J. and Medvedev, O. N. (2020). Applying
37 Generalizability Theory to Differentiate Between Trait and State in the Five Facet Mindfulness
38 Questionnaire (FFMQ). *Mindfulness*, *11*(4), 953-963. doi:10.1007/s12671-020-01324-7
39
40
41

42
43 Vispoel, W. P., Morris, C. A. and Kilinc, M. (2018). Applications of generalizability theory and
44 their relations to classical test theory and structural equation modeling. *Psychological Methods*,
45 *23*(1), 1.
46
47

48
49 Whimbey, A., Vaughan, G. M. and Tatsuoka, M. M. (1967). Fixed effects vs random effects:
50 estimating variance components from mean squares. *Perceptual and Motor Skills*, *25*(2), 668-
51 668.
52
53
54
55
56
57

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Winblad, B., *et al.* (2004). Mild cognitive impairment–beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of internal medicine*, 256(3), 240-246.

For Review Only

Figure/Table legend list**Figure 1**

CONSORT diagram for participants who completed cognitive assessments inclusive of dementia-diagnosed cases at each wave/occasion.

Table 1

G-and D-study estimates for the MAC-Q including standard errors of the grand mean (SE), Coefficient G relative (Gr), Coefficient G absolute (Ga), Trait Component Index (TCI), State Component Index (SCI), grand mean (GM), variance components (in %), and for the Person (P) × Occasion (O) × Item (I) design including interactions with subtracting one item at a time (n = 232).

Table 2

G- and D-study reliability estimates and variance components for the Person (P) × Occasion (O) × Item (I) design including interactions for the Total IQCODE and after removing items with SCI ≥ 0.40 , and occasion.

Table 3

Logistic Regression Model Coefficients for the MAC-Q and IQCODE variables across occasions 2-5 predicting the incidence of dementia.

Table 4

Linear Regression Model Coefficients for the MAC-Q and IQCODE variables across occasions 1-4 predicting the global cognition scores

Figure 1

CONSORT diagram for participants who completed cognitive assessments inclusive of dementia-diagnosed cases at each wave/occasion.

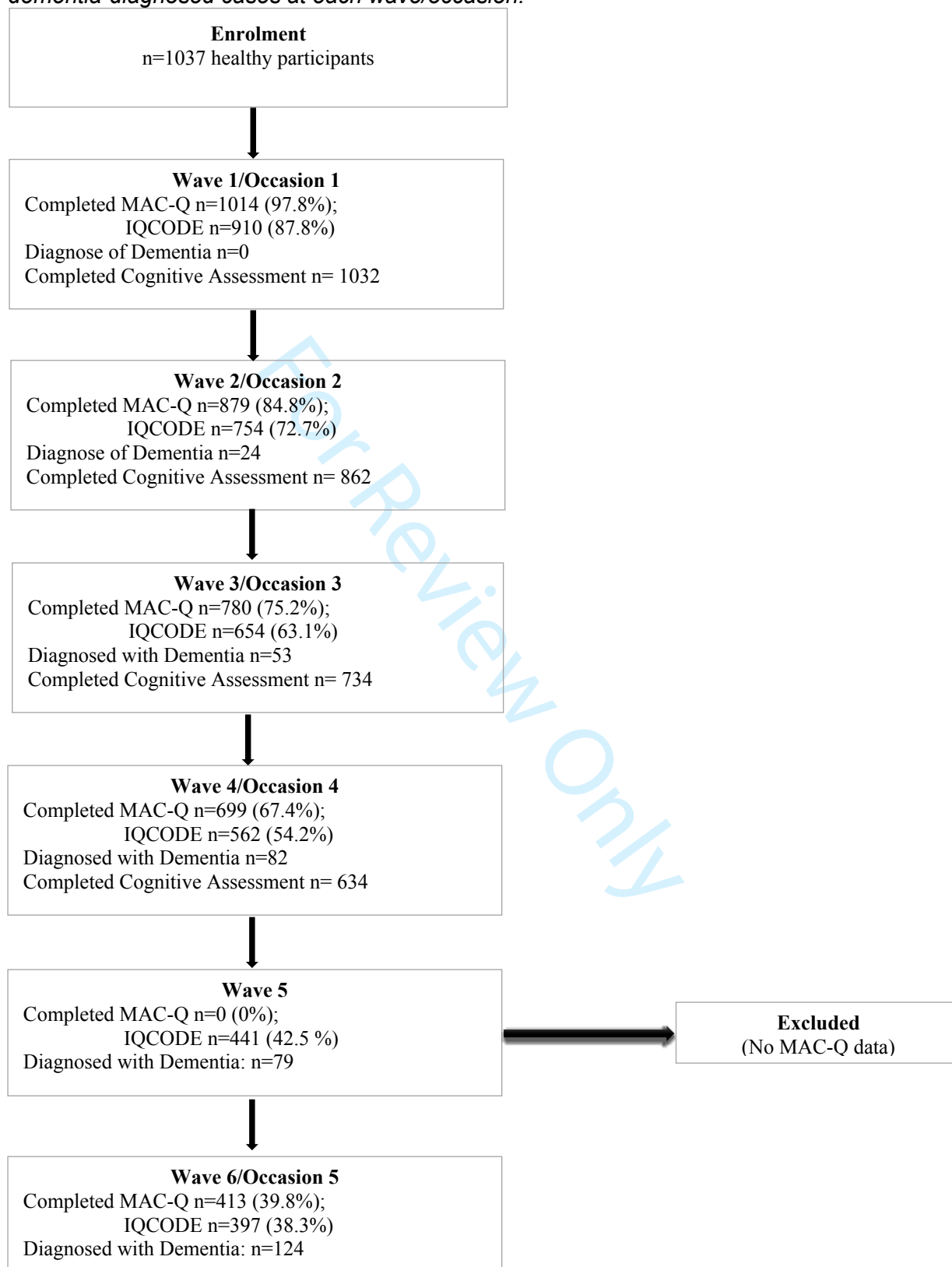


Table 1

G-study estimates for the MAC-Q and IQCODE including standard errors of the grand mean (SE), Coefficient G relative (*Gr*), Coefficient G absolute (*Ga*), Trait Component Index (TCI), State Component Index (SCI), grand mean (GM), variance components (in %), and for the Person (*P*) × Occasion (*O*) × Item (*I*) design including interactions (*n* = 232).

Facets	MAC-Q		IQCODE	
	σ^2	%	σ^2	%
P	0.044		0.019	
I	0.000	2.9	0.000	2.2
O	0.001	5.5	0.000	2.1
PxI	0.005	35.7	0.001	39.6
PxO	0.002	12.6	0.001	37.5
IxO	0.000	3.5	0.000	1.5
PxIxO	0.005	39.7	0.001	17.1
Grand mean	3.224		3.155	
SE	0.042		0.137	
<i>Gr</i>	0.80		0.86	
<i>Ga</i>	0.77		0.86	
TCI	0.96		0.95	
SCI	0.04		0.05	

Note: Numbers in bold signify acceptable reliability/generalizability coefficients

Table 2

Logistic Regression Model Coefficients for the MAC-Q and IQCODE variables across occasions 2-5 predicting the incidence of dementia.

Predicting Diagnosis	MPC	β	SE (β)	p	$Exp(\beta)$ [95% CI]
<u>Occasion 2:</u>					
MAC-Q (5-item)	97.6	1.21	0.13	0.35	1.13 [0.88, 1.45]
MAC-Q (6-item)	96.9	-0.10	0.11	0.37	0.91 [0.73, 1.13]
IQCODE	98.2	5.73	0.90	<0.001	308.80 [53.24, 1791.26]
<u>Occasion 3</u>					
MAC-Q (5-item)	93.9	0.28	0.07	<0.001	1.32 [1.14, 1.53]
MAC-Q (6-item)	94.0	0.30	0.06	<0.001	1.35 [1.19, 1.53]
IQCODE	94.8	4.05	0.53	<0.001	57.13 [20.07, 162.66]
<u>Occasion 4</u>					
MAC-Q (5-item)	90.8	0.29	0.06	<0.001	1.34 [1.20, 1.50]
MAC-Q (6-item)	90.9	0.25	0.05	<0.001	1.29 [1.17, 1.42]
IQCODE	92.3	3.58	0.39	<0.001	36.01 [16.74, 77.48]
<u>Occasion 5</u>					
MAC-Q (5-item)	80.8	0.16	0.07	0.02	1.17 [1.03, 1.33]
MAC-Q (6-item)	78.2	0.05	0.30	0.30	1.05 [0.96, 1.14]
IQCODE	85.3	3.65	0.41	<0.001	38.60 [17.31, 86.108]

Note: MPC: Model Percentage Correct; $Exp(\beta)$: the exponentiation of the β coefficient.

Table 3

Linear Regression Model Coefficients for the MAC-Q and IQCODE variables across occasions 1-4 predicting the global cognition scores.

Predicting Cognition	R^2	β	$SE(\beta)$	p	Standardised β [95% CI]
<u>Occasion 1</u>					
MAC-Q (5-item)	0.002	-0.04	0.03	0.13	-0.05 [-0.11, 0.01]
MAC-Q (6-item)	0.004	-0.04	0.02	0.06	-0.06 [0.10, 1.16]
IQCODE	0.030	-1.02	0.19	<0.001	-0.17 [-0.24, -0.11]
<u>Occasion 2</u>					
MAC-Q (5-item)	0.022	-0.14	0.03	<0.001	-0.15 [-0.22, -0.08]
MAC-Q (6-item)	0.012	-0.08	0.03	<0.01	-0.11 [-0.18, -0.04]
IQCODE	0.055	-1.31	0.20	<0.001	-0.24 [-0.31, -0.17]
<u>Occasion 3</u>					
MAC-Q (5-item)	0.011	-0.10	0.04	<0.01	-0.11 [-0.18, -0.03]
MAC-Q (6-item)	0.014	-0.09	0.03	0.00	-0.12 [-0.18, -0.05]
IQCODE	0.040	-0.94	0.19	<0.001	-0.20 [-0.28, -0.12]
<u>Occasion 4</u>					
MAC-Q (5-item)	0.050	-0.21	0.04	<0.001	-0.23 [-0.31, -0.15]
MAC-Q (6-item)	0.047	-0.16	0.03	0.00	-0.22 [-0.29, -0.14]
IQCODE	0.085	-1.12	0.16	<0.001	-0.29 [-0.37, -0.21]

Note: $SE(\beta)$: standard error of the β coefficient.