Published in final edited form as:

Alzheimer Dis Assoc Disord. 2012 October ; 26(4): 344–351. doi:10.1097/WAD.0b013e3182420bdf.

Subjective cognitive complaints of older adults at the population level: An item response theory analysis

Beth E. Snitz, PhD¹, Lan Yu, PhD², Paul K. Crane, MD, MPH³, Chung-Chou H. Chang, PhD⁴, Tiffany F. Hughes, PhD, MPH², and Mary Ganguli, MD, MPH^{2,1,5}

¹Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA

²Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA

³Division of Internal Medicine, University of Washington, Seattle, WA, USA

⁴Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA

⁵Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

Subjective cognitive complaints (SCCs) are increasingly a focus in studies of prodromal Alzheimer disease (AD) and risk for dementia. Little is known about the optimal approach to measure SCCs. We used item response theory (IRT) to examine characteristics of 24 SCC items in a sample of 3,495 older adults pooled from four community-based studies. We investigated the potential advantages of IRT scoring over conventional scoring, based on participants' item response patterns. Items most likely endorsed by individuals low in SCC severity relate to word retrieval and general subjective memory decline. Items likely endorsed only by individuals high in SCC severity relate to non-episodic memory changes, such as decline in comprehension, judgment and executive functions, praxis and procedural memory, and social behavior changes. IRT scoring of SCCs was associated with performance on objective cognitive test performance above and beyond total SCC scores, and was associated with objective cognitive test performance among participants endorsing only one SCC item. Thus, IRT scoring captures additional information beyond a simple sum of SCC symptoms. Modern psychometric approaches including IRT may be useful in developing 1) brief community screening questionnaires, and 2) more sensitive measures of very subtle subjective decline for use in prodromal AD research.

Keywords

Subjective memory; item response theory; dementia; neuropsychological tests; subjective cognitive impairment

Introduction

In recent years, subjective cognitive complaints (SCCs) or concerns have received increased attention in cognitive aging research. Subjective memory complaints were an initial criterion

Address correspondence to: Beth Snitz, PhD, Department of Neurology, University of Pittsburgh, 3600 Forbes Avenue, Suite 8062, Pittsburgh, PA 15213. Phone: 412-692-4820. Fax: 412-864-2405. snitzbe@upmc.edu.

Conflicts of interest: None.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

for the syndrome of Mild Cognitive Impairment (MCI),¹ with a revision broadening the criterion to include general cognitive complaints in domains other than memory.² SCCs are increasingly investigated as a predictor of cognitive decline or dementia in aging.³⁻⁵ Studies of self-referring individuals in specialty memory clinic settings may offer the clearest evidence that subjective cognitive impairment, even in the absence of objective cognitive deficits, may serve as a marker of future decline⁵ or of presence of Alzheimer disease (AD) biomarkers.⁶⁻⁸ Even population studies, in which “complaints” are uniformly elicited by standardized questionnaires, show that people with cognitive complaints are at greater risk of progressing to dementia than those without cognitive complaints.^{4, 9}

Cognitive self-appraisal has practical advantages of simplicity and cost-effectiveness compared to formal cognitive testing and biomarker evaluation, and may be well-suited for large-scale community screening. However, there is currently no consensus about how to optimally assess subjective complaints, especially in population settings where the underlying rates of complaints are lower than in clinical settings.¹⁰ Studies which elicit complaints have used a variety of questionnaires or scales, many assessing SCCs with a few items or even a single question. Not surprisingly, findings of predictive utility from community and population studies are variable and inconsistent.^{11, 12} Psychometric properties of SCC items have rarely been investigated;¹³ little is known about which questions to ask, about information yielded by different types of questions/items, and about the quality of measurement.

Item response theory (IRT) comprises a group of psychometric modeling techniques for the analysis of item-level data from self-report questionnaires or cognitive tests. Originally developed in educational achievement testing settings, IRT has been increasingly used in health outcome measurement. Two large NIH initiatives currently use measurement science to improve assessment of self-reported health outcomes (PROMIS)¹⁴ and neurological and behavioral function (NIH Toolbox¹⁵). Tools from IRT can be used to investigate psychometric properties of existing scales; refine or shorten scales (reducing patient burden); develop new scales; and score respondents on an underlying latent trait (or ability) and precisely measure interindividual variation. Unlike classical test theory (CTT), IRT measures item difficulty and person trait level on the same metric, allowing them to be meaningfully compared. Other advantages of IRT include linear (interval) scaling; test precision (reliability) measured over the complete range of the trait/ability; and invariance of IRT model parameters, which allows “linking” of different scales measuring the same construct and implementation of computerized adaptive testing (CAT).¹⁶

Previous applications of IRT in health research include self-reported physical functioning,¹⁷ sleep disturbance,¹⁸ quality of life,¹⁹ and pain.²⁰ IRT approaches in cognitive aging research include refined measurement of neuropsychological domain performance,^{21, 22} dementia screening instruments,^{23, 24} and depressive symptoms.²⁵ Use of IRT to examine cognitive item bias in ethnically diverse older populations has also been reported.^{22, 26} To our knowledge, no previous studies have used IRT approaches to evaluate subjective cognitive complaints.

We are using an IRT approach to address the overall research question of whether SCCs are a valid marker of cognitive decline in aging. The aims of the present methodological study were to 1) conduct an IRT analysis on subjective cognitive complaint data in a population study of older adults; 2) compare psychometric properties of individual SCC items; and 3) compare IRT-based scoring of individuals to traditional scale scoring, particularly regarding associations between SCCs and cognitive test performance.

Methods

Overview

Methods involved three main steps: a) factor analysis of 27 common SCC items in a pooled community sample (n=3495) to assess whether SCC items met assumptions made by IRT models; b) IRT item calibration in the same pooled sample; and c) IRT scoring of individuals from the largest of the studies, the Monongahela-Youghiogheny Healthy Aging Team (MYHAT) Project.

Sample description

To maximize precision of IRT item parameter estimates and to achieve the broadest coverage of the range of SCC severity, we sought the largest sample possible of 27 common SCC items pooled from four community studies of cognition in aging in Southwestern Pennsylvania. Studies included: 1) the MYHAT Project,²⁷ a prospective population study of MCI in small towns (baseline data, n=1,982); 2) the Steel Valley Seniors Survey (SVSS),²⁸ a study of dementia incidence and outcomes in an elderly general practice patient population (n=358); 3) Monongahela Independent Elders Survey (MoVIES),²⁹ an epidemiologic community-based study of aging and dementia in rural Southwestern Pennsylvania (wave 6 data, n=651); and 4) a study of physician behavior and cognitive outcomes in an older adult primary care patient population³⁰ (baseline data, n=534). All participants provided written informed consent according to protocols approved by the University of Pittsburgh IRB. Participants (n=30) were excluded from analyses only if they had missing data on all the SCC items of interest, leaving 3,495 individuals in the pooled item calibration sample.

SCC items

Standardized SCC items were originally developed for the MoVIES study and have been adapted to the other studies.³⁰ We identified a set of 27 items with similar response formats across the studies that each item had a minimum sample size of 1,000 respondents. Items overlapped to various degrees across the studies. Fourteen items were common to all four studies; 4 items were common to three studies; 9 items were unique to one study (MYHAT). Twenty-one items had the common prompt and response choices: “*Do you think you are the same, better or worse than you used to be at...*” We categorized these response choices as “same or better” vs. “worse” for the current analyses. Item content was weighted toward memory-related changes (e.g., remembering less well than one year ago), but also included non-memory items, such as those relating to changes in day-to-day executive functioning and judgment (e.g., handling household emergencies, solving problems), and language (understanding what is being said, word-finding) (Table 2).

Background to IRT modeling

IRT models describe the probability of endorsing an item as a function of a latent trait of interest. By convention, the latent trait is scaled along a dimension called theta. Parametric IRT models identify one or more parameters that describe the relationship between theta and item responses. Item parameters include item *difficulty* (*b*) (i.e., at what value of theta an individual has a 0.50 probability of endorsing an item) and *discrimination* (*a*) (i.e., how well an item distinguishes among individuals with theta (trait) levels close to its level of difficulty). In this study, the latent trait is subjective cognitive complaints. IRT provides psychometric information regarding each SCC item separately, as well as psychometric information for the full scale. Additionally, IRT provides theta estimates for individual participants. In this way, an individual's responses can be used to estimate his or her SCC severity relative to the rest of the sample population.

Step 1: Factor analysis

An important assumption of IRT models is *unidimensionality* of the latent trait of interest (i.e., SCC severity). We used factor analysis to determine if SCC items fit sufficiently well to a one-factor model to meet this assumption for IRT analysis. We randomly split the pooled multi-study data sample (n=3495) and conducted an exploratory factor analysis (EFA) using Mplus v. 6.1 with robust weighted least squares estimation and Geomin oblique rotation in one of the sub-samples (n=1747). Ratio of 1st to 2nd eigenvalues in excess of 4:1 was used as a criterion for support of the unidimensionality assumption³¹. We then confirmed the factor structure in an independent confirmatory factor analysis (CFA) on the second sub-sample (n=1748). Standardized factor loadings of 0.4 or higher was a criterion for item retention in both the EFA and the CFA. Criteria for fit indices were < 0.08 for RMSEA and > 0.90 for the comparative fit index (CFI) and the Tucker-Lewis index (TLI). In addition, we ran two alternative CFA models: a one-factor model with selected item-pair residual covariances included, and a theoretically-derived bifactor model.³¹ For each alternative confirmatory model, factor loadings on the primary factor were compared to the simple unidimensional model, and the correlations between factor scores were computed.

Step 2: IRT item calibration

IRT parameters can be visualized by plotting item characteristic curves (ICCs). ICCs plot the probability of a response (endorsing a symptom) given the level of the underlying characteristic measured by the whole scale (such as SCC severity). In this study, ICCs are defined by difficulty and discrimination, described above, which govern the shape and position of S-shaped curves. Item difficulty, the *b* parameter, is so termed from educational testing settings, and is alternatively called item *location*. The *b*-parameter governs the position of the ICC along the x-axis and reflects the point on the trait dimension at which a person has a 50% probability of endorsing the item. In clinical and epidemiological studies, item location / difficulty also reflects the prevalence of symptoms. If the *b*-parameter is low, it indicates the item (or symptom) is frequently endorsed even among low-trait individuals. In contrast, high *b*-parameters indicate the item is likely to be endorsed only among high-trait individuals (e.g., with more severe subjective cognitive complaints). Item discrimination, the *a*-parameter, governs the steepness of the slope of the ICC at the inflection point of the S-shaped curve. A lower *a*-parameter is associated with a more gradual slope; items with higher *a*-parameters have steeper curves and therefore discriminate more finely among respondents with theta values close to the value of the item's *b*-parameter.

Item parameter estimates were derived from the final item set in the total pooled multi-study sample (n=3,495) using Multilog v. 7.0.3 with a 2 parameter logistic (2PL) model and marginal maximal likelihood estimation procedure.

Step 3: IRT theta scores

MYHAT study participants at baseline were scored along the theta dimension, reflecting the underlying SCC trait based on individual item responses (Multilog v. 7.0.3, maximum a posterior scoring). Participants with complete SCC data were selected (n=1,789) because the analyses of standard scores (see below) required complete data. By convention, theta scores represent standardized latent trait values centered on a mean of 0 and standard deviation of 1. Higher theta scores indicate higher trait levels (e.g., SCC severity). Comparisons between IRT theta scoring and traditional scale scoring (i.e., a sum of the number of symptoms endorsed) were examined via associations with demographic variables, and with concurrent objective cognitive test performance. Finally, sensitivity of theta scores to cognitive test performance for a fixed level of total score was examined, to explore the potential utility of theta scoring. The rationale for these analyses is that if the total score were a good reflection

of the underlying trait measured by the test, there should be no systematic relationship between IRT scores and objective cognitive functioning among people with a particular total score. We selected total SCC scores of 1 and 5 *a priori* because they represented low and moderate degrees of symptom endorsement, respectively, and each yielded sample sizes > 100.

Cognitive test performance

We ran a series of regression models at each level of total score, with theta scores predicting cognitive test performance. Alpha was set to $p < .10$ for these exploratory analyses. Cognitive measures included a global cognitive test – the Mini-Mental State Examination (MMSE);³² two memory tests - WMS-R Logical Memory and Visual Reproduction³³ and the Fuld Object Memory Evaluation (OME);³⁴ three tests of psychomotor speed and executive functioning - Trail Making Test A & B³⁵ and clock drawing; two tests of productive language – phonemic (FAS) and semantic (animal) fluency³⁵; and one test of visuospatial construction - WAIS-III Block Design.³⁶ Depressive symptoms were measured by the modified Center for Epidemiologic Studies Depression (mCESD) scale.^{37,38} Functional ratings were measured with the Clinical Dementia Rating (CDR) scale.

Results

Descriptive statistics

Table 1 summarizes demographic and clinical characteristics of the pooled multi-study item calibration sample. While the majority of respondents had MMSE scores within the normal range, a sizable proportion had impaired MMSE scores (> 16% less than MMSE 25). Nearly 4% had CDR scores of 1 or greater, indicating dementia based on functional ratings, while 28.2% had a CDR score of 0.5, reflecting “questionable dementia” or possible mild cognitive impairment. The sample thus included participants with a broad range of cognitive functioning, the majority being rated as normal (CDR=0).

Dimensionality evaluations

Table 2 presents frequencies of item responses. An EFA on the original 27 SCC items yielded factor loadings on the 1-factor solution ranging from approximately 0.4 to 0.8 for all items but two with loadings near 0. These two items were dropped (*Did your memory get worse suddenly or gradually? How long ago did you notice change in your memory?*). The independent CFA of a 1-factor model on the remaining 25 items yielded all items with factor loadings > 0.4 except one item (*Is remembering less well a consistent problem?*), which was dropped. The resulting 1-factor CFA on 24 items yielded the following fit indices: RMSEA = 0.029 (90% CI 0.026 to 0.031); CFI = 0.934; TLI = 0.928. An EFA in the split-half sample on these 24 items yielded a first eigenvalue of 9.207 followed a second eigenvalue of 2.011 (ratio 1:2 = 4.6). Supplemental Table 1 provides CFA factor loadings for these final 24 SCC items. Supplemental Methods provides further detail regarding unidimensionality assessment. Based on all of these analyses, we concluded that the remaining SCC items were sufficiently unidimensional to use IRT.³¹

IRT item parameter estimates

Table 2 also presents item difficulty (*b*) and discrimination (*a*) parameters, in order of item difficulty. The item with the lowest difficulty parameter, or “easiest” to endorse, was *Remembering names of people you met only recently* followed by *Do you feel you remember things less well than a year ago*. The item with the highest difficulty parameter was *Worse at getting along with people, talking or behaving the way you used to*, likely endorsed only by high-trait individuals. Those items with high discrimination (*a*) parameters

(e.g., *Worse at handling a household emergency; Worse at remembering a recent major event; Worse at understanding what is going on around you*) are the most efficient at differentiating individuals on the underlying trait of SCC severity at their respective b-parameter locations.

For illustrative purposes, item characteristic curves for three selected items are presented in Figures 1-3. As described above, ICCs graphically depict the probability of an item being endorsed (y-axis) against the level of the underlying trait (x-axis). The left-right location of the inflection point of the curve is determined by the difficulty parameter. The more people endorse an item, the further to the left the curve will be, as it is “easier” to endorse that item. Items that are rarely endorsed will be further to the right, as they are more “difficult” to endorse. The items *Remembering a recent major event* (Figure 1) and *Remembering people and events from long ago* (Figure 2) have relatively high difficulty parameters, meaning they are likely only endorsed by individuals with high levels of the underlying trait. In contrast, the ICCs for *Remembering names of people you met only recently* (Figure 3) is shifted to the left, indicating likely endorsement by persons lower on the underlying trait (as well as higher). A steep slope of the curve at the point of inflection (i.e., a high *a*-parameter) indicates a sharp rise in probability of item endorsement, e.g., Figure 1, as the level of underlying trait increases only slightly. The item *Remembering people and events from long ago* (Figure 2) is less precise in differentiating individuals with underlying ability levels close to its difficulty level as the slope of the curve is shallower. For each figure, the frequencies of item endorsement as a function of the underlying trait are shown above the ICCs, to illustrate the correspondence between item parameters and the ICCs.

IRT total scale information

Information for the total scale, analogous to scale reliability in classical test theory, ranged from 1.06 (for $\theta = -3.0$) to 18.4 (for $\theta = 2.0$). The peak range of information (15.0 to 18.3) was for the θ range of 1.4 to 2.6, indicating the scale as a whole is most reliable at measuring high levels of underlying SCCs. Test information of 20 corresponds approximately with conventional reliability of 0.95, and test information of 10 corresponds approximately with conventional reliability of 0.90. There is an inverse numerical relationship between information and the standard error of measurement, (see Supplemental Figure).

SCCs in the MYHAT study: Comparison of IRT scoring and traditional scoring

Recruitment and demographic details of the MYHAT study baseline cohort have been reported.²⁷ Briefly, after initial age-stratified random sampling by voter registration records, $n=1,982$ participants age ≥ 65 passed an initial screening stage with an age-education corrected MMSE ≥ 21 . Mean (SD) age was 77.6 (7.4) years; 61.1% were women. The median educational level was high school graduate; 94.8% were white, largely representative of older adults in this small-town region south of Pittsburgh. Apolipoprotein E (*APOE*) genotyping was available on $n=1778$; 18.8% ($n=372$) were *APOE* $\epsilon 4$ allele carriers.

The same 24 SCC item set was used to compare IRT θ scoring with the traditional total score (i.e., sum of endorsed symptoms). Using the traditional total score approach, the modal number of SCC symptoms endorsed at baseline was 0. Seventy percent of the cohort endorsed 3 items or fewer. The IRT θ scores ranged from -0.83 (lowest SCCs) to 2.69 (highest SCCs); with a mean θ of 0.01 (SD 0.73). Figure 4 illustrates the relationship between traditional total scores, on the y-axis, and θ scores, which continue to vary at almost all levels of total score.

Associations between SCCs, demographic and depressive symptoms—Total SCC score and theta score were highly correlated ($r=0.95$, $p < .001$). They showed comparable magnitude of correlation with age (total score $r=0.12$; theta $r=0.15$, $p's < .001$) and with education (total score $r=-0.07$; theta $r=-0.08$, $p's < .001$), indicating increasing SCC severity with increasing age and with lower years of formal education. Neither differed by gender ($p's > .15$). Neither total score nor theta score were associated with presence of one or more copies of the *APOE* $\epsilon 4$ allele ($p's > .30$). Both scores were similarly associated with mCESD score (total score $r=0.25$; theta $r=0.23$; $p's < .001$), indicating higher SCCs with increasing depressive symptoms.

Associations between SCCs and cognitive performance—Total SCC score and theta score were highly similarly and significantly associated ($p's < .001$) with cognitive test performance on all measures in the expected direction (higher SCCs associated with worse test performance): MMSE (total score $r=-0.19$; theta $r=-0.19$); WMS-R Logical Memory II (total score $r=-0.14$; theta $r=-0.14$); WMS-R Visual Reproduction (total score $r=-0.14$; theta $r=-0.14$); Fuld OME (total score $r=-0.21$; theta $r=-0.21$); Trails A time (total score $r=0.14$; theta $r=0.15$); Trails B time (total score $r=0.18$; theta $r=0.18$); clock drawing (total score $r=-0.13$; theta $r=-0.14$); letter fluency (total score $r=-0.10$; theta $r=-0.10$); semantic fluency (total score $r=-0.19$; theta $r=-0.19$) and WAIS-III Block Design (total score $r=-0.20$; theta $r=-0.20$).

Association of theta scores with cognitive performance when SCC total score is held constant—To explore the potential utility of theta scores above and beyond the traditional total score approach, we selected participants with the same total score and evaluated associations between theta scores and cognitive test performance. Table 3 shows regression model results for theta scores predicting cognitive test performance at fixed SCC total score levels of 1 and 5. For SCC total score = 1, higher theta scores predicted worse performance on tests of mental status, memory and psychomotor speed; at SCC total score = 5, higher theta scores predicted worse performance on tests of executive functions, language and psychomotor speed. Nearly all of the regression estimates indicated associations between theta and cognitive test performance in the predicted direction (i.e., higher theta predicting worse test performance).

Discussion

We have illustrated how IRT, an approach from measurement science, can enhance the assessment of subjective cognitive complaints, an important patient-reported variable in cognitive aging and dementia research. While IRT may be used to develop new tests or questionnaires, here we applied IRT models retrospectively to existing questionnaire data, without altering data collection procedures, to further understand and refine measurement of SCCs in an ongoing population study.

IRT analysis can inform clinicians' and investigators' selection of individual questions to query and provide insight regarding how each item contributes to measurement of the underlying trait. In the present item set, items most likely to be endorsed by individuals *low* on the underlying SCC trait are questions about remembering names, word-finding, and remembering less well compared to the preceding year. Items most likely to be endorsed only by people *high* on the SCC trait address changes in behavior and language, and procedural memory/praxis. This ordering of items by difficulty corresponds to the known spectrum of cognitive changes from normal aging to dementia. Items found to be of low difficulty relate to commonly reported cognitive complaints in normal aging and MCI, e.g., retrieval difficulty of names and other words, and general worsening in memory. In contrast, higher difficulty items correspond to symptoms more characteristic of dementia, with

changes beyond episodic memory to non-memory domains (e.g., language, judgment and executive functioning, praxis, and overt interpersonal difficulties). Since items and persons are measured on the same scale, item characteristics relate directly to trait scoring. Individuals who endorse only low difficulty items will have low IRT SCC scores, while those who endorse some high difficulty items, reflecting changes in cognitive domains beyond episodic memory, will have higher IRT SCC scores.

IRT scores were very highly correlated with standard total scores and the two scores had very similar strengths of associations with demographic variables and with cognitive test performance. However, IRT scoring also uses information about the “path taken” to arrive at a particular total score. That is, at fixed values of the total score, IRT scoring continued to vary meaningfully so that subjective cognitive complaints predicted performance on tests of mental status, memory, language, executive functions and psychomotor speed. Strikingly, even among participants who endorsed *only one SCC item*, IRT scoring of SCCs predicted performance on several objective cognitive tests. Thus, IRT captures additional information beyond the total score. IRT scoring may be particularly well suited to population studies where the base-rates of complaints are low, and finer-grained measurement of individual differences would prove useful.

Item level information gained from an IRT analysis can be used to shorten our SCC scale in the future (i.e., to minimize time burden) by eliminating less informative items, or to optimize the scale for different levels of the underlying trait (e.g., as a brief screen for AD in community settings). Additionally, the IRT parameters facilitate the possibility of administering these items using computerized adaptive testing (CAT).³⁹ In CAT, each survey is individualized so as to only ask the most informative questions for a given respondent (i.e., those items with difficulty parameters near the respondent's latent trait level, as estimated from previous item responses). In this way, only the least number of items needed to achieve a predetermined standard of score precision are administered.

In sum, we have shown that IRT provides better understanding of SCC questionnaire data than total scoring, and that IRT can identify meaningful differences between individuals with the same SCC total score. Future work will investigate potential gain provided by IRT over total scores in predictive validity of SCC for cognitive decline and progression to AD and other dementias. This aim is consistent with recently published recommendations by the NIA/AA workgroup on defining preclinical stages of AD, which note that measures sensitive to very subtle cognitive decline are needed to further realize the potential of subjective and cognitive assessments.⁴⁰ Leveraging approaches from measurement science may significantly advance detection of very subtle cognitive decline, which in turn may further our understanding of how aging- and disease-related biomarkers relate to cognition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by grants from the National Institute on Aging (K23AG038479, R01 AG07562, K24 AG022035).

References

1. Petersen RC, Smith GE, Waring SC, et al. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol*. 1999; 56(3):303–308. [PubMed: 10190820]

2. Winblad B, Palmer K, Kivipelto M, et al. Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med*. 2004; 256(3):240–246.
3. Reisberg B, Gauthier S. Current evidence for subjective cognitive impairment (SCI) as the pre-mild cognitive impairment (MCI) stage of subsequently manifest Alzheimer's disease. *Int Psychogeriatr*. Feb; 2008 20(1):1–16. [PubMed: 18072981]
4. Jessen F, Wiese B, Bachmann C, et al. Prediction of dementia by subjective memory impairment: effects of severity and temporal association with cognitive impairment. *Arch Gen Psychiatry*. 2010; 67(4):414–422. [PubMed: 20368517]
5. Reisberg B, Shulman MB, Torossian C, et al. Outcome over seven years of healthy adults with and without subjective cognitive impairment. *Alzheimer's & Dementia*. Jan; 2010 6(1):11–24.
6. Barnes LL, Schneider JA, Boyle PA, et al. Memory complaints are related to Alzheimer disease pathology in older persons. *Neurology*. Nov 14; 2006 67(9):1581–1585. [PubMed: 17101887]
7. Saykin AJ, Wishart HA, Rabin LA, et al. Older adults with cognitive complaints show brain atrophy similar to that of amnesic MCI. *Neurology*. Sep 12; 2006 67(5):834–842. [PubMed: 16966547]
8. Mosconi L, De Santi S, Brys M, et al. Hypometabolism and Altered Cerebrospinal Fluid Markers in Normal Apolipoprotein E E4 Carriers with Subjective Memory Complaints. *Biol Psychiatry*. 2008; 63(6):609–618. [PubMed: 17720148]
9. Wang L, van Belle G, Crane PK, et al. Subjective memory deterioration and future dementia in people aged 65 and older. *J Am Geriatr Soc*. Dec; 2004 52(12):2045–2051. see comment. [PubMed: 15571540]
10. Jonker C, Launer LJ, Hooijer C, et al. Memory complaints and memory impairment in older individuals. *J Am Geriatr Soc*. Jan; 1996 44(1):44–49. see comment. [PubMed: 8537589]
11. Smith GE, Petersen RC, Ivnik RJ, et al. Subjective memory complaints, psychological distress, and longitudinal change in objective memory performance. *Psychol Aging*. Jun; 1996 11(2):272–279. [PubMed: 8795055]
12. Dik MG, Jonker C, Comijs HC, et al. Memory complaints and APOE-epsilon4 accelerate cognitive decline in cognitively normal elderly. *Neurology*. Dec 26; 2001 57(12):2217–2222. [PubMed: 11756600]
13. Reid LM, MacLullich AMJ. Subjective memory complaints and cognitive impairment in older people. *Dement Geriatr Cogn Disord*. 2006; 22(5-6):471–485. [PubMed: 17047326]
14. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010; 63(11):1179–1194. [PubMed: 20685078]
15. Gershon RC, Cella D, Fox NA, et al. Assessment of neurological and behavioural function: the NIH Toolbox. *Lancet Neurology*. Feb; 2010 9(2):138–139. [PubMed: 20129161]
16. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*. 2007; 16(Suppl 1):5–18. [PubMed: 17375372]
17. Martin M, Kosinski M, Bjorner JB, et al. Item response theory methods can improve the measurement of physical function by combining the modified health assessment questionnaire and the SF-36 physical function scale. *Qual Life Res*. May; 2007 16(4):647–660. [PubMed: 17334829]
18. Buysse DJ, Yu L, Moul DE, et al. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. *Sleep*. Jun 1; 2010 33(6):781–792. [PubMed: 20550019]
19. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res*. Aug; 1997 6(6):595–600. [PubMed: 9330558]
20. Amtmann D, Cook KF, Jensen MP, et al. Development of a PROMIS item bank to measure pain interference. *Pain*. Jul; 2010 150(1):173–182. [PubMed: 20554116]
21. Mungas D, Reed BR, Kramer JH, et al. Psychometrically matched measures of global cognition, memory, and executive function for assessment of cognitive decline in older persons. *Neuropsychology*. Jul; 2003 17(3):380–392. [PubMed: 12959504]

22. Mungas D, Reed BR, Crane PK, et al. Spanish and English Neuropsychological Assessment Scales (SENAS): Further Development and Psychometric Characteristics. *Psychological Assessment*. Dec; 2004 16(4):347–359. [PubMed: 15584794]
23. Crane PK, Narasimhalu K, Gibbons LE, et al. Item response theory facilitated cocalibrating cognitive tests and reduced bias in estimated rates of decline. *J Clin Epidemiol*. Oct; 2008 61(10): 1018–1027.e1019. [PubMed: 18455909]
24. Gibbons LE, McCurry S, Rhoads K, et al. Japanese-English language equivalence of the Cognitive Abilities Screening Instrument among Japanese-Americans. *Int Psychogeriatr*. Feb; 2009 21(1): 129–137. [PubMed: 18947456]
25. Jones RN, Fonda SJ. Use of an IRT-based latent variable model to link different forms of the CES-D from the Health and Retirement Study. *Soc Psychiatry Psychiatr Epidemiol*. Oct; 2004 39(10): 828–835. [PubMed: 15669664]
26. Crane PK, Gibbons LE, Jolley L, et al. Differential item functioning related to education and age in the Italian version of the Mini-mental State Examination. *Int Psychogeriatr*. Sep; 2006 18(3):505–515. [PubMed: 16478571]
27. Ganguli M, Snitz B, Vander Bilt J, et al. How much do depressive symptoms affect cognition at the population level? The Monongahela-Youghiogheny Healthy Aging Team (MYHAT) study. *Int J Geriatr Psychiatry*. Nov; 2009 24(11):1277–1284. [PubMed: 19340894]
28. Ganguli M, Rodriguez E, Mulsant B, et al. Detection and management of cognitive impairment in primary care: The Steel Valley Seniors Survey. *J Am Geriatr Soc*. Oct; 2004 52(10):1668–1675. [PubMed: 15450043]
29. Ganguli M, Dodge HH, Chen P, et al. Ten-year incidence of dementia in a rural elderly US community population: the MoVIES Project. *Neurology*. Mar 14; 2000 54(5):1109–1116. [PubMed: 10720283]
30. Snitz BE, Morrow LA, Rodriguez EG, et al. Subjective memory complaints and concurrent memory performance in older patients of primary care providers. *J Int Neuropsychol Soc*. Nov; 2008 14(6):1004–1013. [PubMed: 18954480]
31. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* May. 2007; 45(5 Suppl 1):S22–31.
32. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. Nov; 1975 12(3):189–198. [PubMed: 1202204]
33. Wechsler, D. Wechsler Memory Scale Revised: The Psychological Corporation. 1987.
34. Fuld, PA. Fuld Object-Memory Evaluation. Woodale, IL: Stoelting Company; 1981.
35. Spreen, O.; Strauss, E. A Compendium of Neuropsychological Tests. New York: Oxford University Press; 1998.
36. Wechsler, D. Wechsler Adult Intelligence Scale. 3rd. San Antonio, TX: The Psychological Corporation; 1997.
37. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*. 1977; 1:385–401.
38. Ganguli M, Gilby J, Seaberg E, et al. Depressive symptoms and associated factors in a rural elderly population: the MoVIES Project. *Am J Geriatr Psychiatry*. 1995; 3:144–160.
39. Cella D, Gershon R, Lai JS, et al. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Qual Life Res*. 2007; 16(Suppl 1):133–141. [PubMed: 17401637]
40. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*. 7(3): 280–292.

Figure 1A & B.

High difficulty – high discrimination item (*19. Worse at remembering a major recent event*). This item is more likely to be endorsed by high-trait individuals, as indicated by the high b-parameter, and is relatively precise in measuring individual differences among respondents, as indicated by the high a-parameter. A) Increasing frequencies of item endorsements along higher levels of the SCC latent trait. B) Item Characteristic Curve based on IRT modeling.

Figure 2A & B.

High difficulty – low discrimination item (21. *Worse at remembering events and people from long ago*). This item is more likely to be endorsed by high-trait individuals, but it is less precise in measuring individual differences among respondents. A) Increasing frequencies of item endorsements along higher levels of the SCC latent trait. B) Item Characteristic Curve based on IRT modeling.

Figure 3A & B.

Low difficulty – moderate discrimination item (*1. Worse at remembering names of people you only met recently*). This item is relatively “easy” to endorse by lower-trait individuals and it is moderately precise in measuring individual differences among respondents. A) Increasing frequencies of item endorsements along higher levels of the SCC latent trait. B) Item Characteristic Curve based on IRT modeling.

Figure 4.

Scatterplot of two alternative approaches to scoring persons in the MYHAT study (n=1,789): IRT theta (latent trait) scores by traditional total scores (the sum of endorsed symptoms out of 24). IRT theta scoring takes into account individual response patterns and utilizes item-level information to provide a more fine-grained measurement scale. The figure shows that theta continues to vary at most levels of the total score. The two levels of total scores in squares on the y-axis (1 and 5) correspond to analyses summarized in Table 3.

Table 1
Demographic and clinical characteristics of pooled tem calibration sample (n=3,495)

Age (mean, SD, range)	77.6, SD 7.1, 65 - 99
Education (n, %)	
Less than HS	n = 650, 18.6 %
HS / GED	n = 1,520, 43.5 %
Greater than HS	n = 3,424, 37.9 %
Sex (n, % female)	n = 2,184, 62.5%
Race (n, %)	
White	n = 3,361, 95.3 %
Other	n = 164, 4.7 %
MMSE score (n, %)	
8 - 18	n = 38, 1.1 %
19 - 24	n = 532, 15.2 %
25 - 30	n = 2,923, 83.7 %
CDR global score (n, % out of n=2,961 available)	
0	n = 2,014, 68.0 %
0.5	n = 836, 28.2 %
1 or greater	n = 111, 3.7 %

Abbreviations: MMSE = Mini Mental State Examination; CDR = Clinical Dementia Rating scale.

Table 2
Item response frequencies and IRT parameters from item calibration sample (n=3,495),
in order of item difficulty (lowest to highest)

Item	Response choices	Response frequencies (n)	Discrimination (a)	Difficulty (b)
1. Remembering names of people you only met recently?	<i>Worse</i>	33.4% (1165)	1.69	0.58
	<i>Same or better</i>	66.6 % (2322)		
2. Do you feel you remember things less well than you did a year ago?	<i>Yes</i>	30.2% (1053)	1.48	0.76
	<i>No</i>	69.8% (2438)		
3. Finding the right word to use to describe something you know well?	<i>Worse</i>	31.7% (1108)	1.25	0.76
	<i>Same or better</i>	68.3% (2382)		
4. Remembering names of people you've known for a long time?	<i>Worse</i>	26.0% (906)	1.29	1.02
	<i>Same or better</i>	74.0 % (2585)		
5. Remembering things that happened or were said a few days ago?	<i>Worse</i>	21.8% (763)	1.61	1.09
	<i>Same or better</i>	78.1% (2728)		
6. Remembering a few things you wanted to buy in the store, without writing them down?	<i>Worse</i>	16.8% (581)	1.86	1.26
	<i>Same or better</i>	83.2% (2887)		
7. Remembering where you've put things that you use often?	<i>Worse</i>	19.2% (670)	1.40	1.33
	<i>Same or better</i>	80.8% (2821)		
8. Are you worried about problems with remembering?	<i>Yes</i>	20.9% (415)	1.39	1.57
	<i>No</i>	79.1% (1571)		
9. Solving problems as easily as before?	<i>Worse</i>	8.6% (170)	2.23	1.58
	<i>Same or better</i>	91.4% (1807)		
10. Remembering appointments, messages, etc.?	<i>Worse</i>	10.0% (349)	1.95	1.65
	<i>Same or better</i>	90.0% (3141)		
11. Remembering what day/date/month it is?	<i>Worse</i>	10.0% (287)	1.92	1.69
	<i>Same or better</i>	90.0% (2583)		
12. Remembering when events happened in relation to each other?	<i>Worse</i>	7.8% (155)	1.99	1.73
	<i>Same or better</i>	92.2% (1824)		
13. Remembering telephone numbers of people whom you call often?	<i>Worse</i>	9.7% (338)	1.78	1.75
	<i>Same or better</i>	90.3% (3144)		
14. Remembering the details of a major event?	<i>Worse</i>	5.4% (107)	2.44	1.80
	<i>Same or better</i>	94.6% (1868)		
15. Remembering a familiar/favorite recipe without looking it up?	<i>Worse</i>	5.6% (151)	1.93	2.01
	<i>Same or better</i>	94.4% (2528)		
16. Was this enough of a problem that you talked to your doctor/nurse about it?	<i>Yes</i>	10.2% (238)	1.68	2.04
	<i>No</i>	89.8% (2095)		
17. Keeping up with hobbies and interests?	<i>Worse</i>	4.9% (97)	1.91	2.08
	<i>Same or better</i>	95.1% (1882)		
18. Handling a household emergency like a plumbing leak or a kitchen fire?	<i>Worse</i>	1.9% (37)	2.70	2.25
	<i>Same or better</i>	98.1% (1941)		
19. Remembering a recent major event? (e.g., trip, wedding)	<i>Worse</i>	1.7% (33)	2.68	2.31
	<i>Same or better</i>	98.3% (1945)		

Item	Response choices	Response frequencies (n)	Discrimination (a)	Difficulty (b)
20. Remembering to do important routine things like turn off the stove or lock the door?	<i>Worse</i>	4.2% (145)	1.72	2.41
	<i>Same or better</i>	95.8% (3338)		
21. Remembering things (events, people, etc.) from a long time ago?	<i>Worse</i>	10.0% (350)	1.01	2.48
	<i>Same or better</i>	89.9% (3141)		
22. Remembering how to use familiar appliances, tools, gadgets?	<i>Worse</i>	1.3% (44)	2.52	2.60
	<i>Same or better</i>	98.7% (3433)		
23. Understanding what is going on or being said around you?	<i>Worse</i>	0.9% (17)	2.60	2.64
	<i>Same or better</i>	99.1% (1962)		
24. Getting along with people, or talking and behaving the way you used to do?	<i>Worse</i>	1.0% (20)	1.90	3.04
	<i>Same or better</i>	99.0% (1959)		

Table 3
Regression model results for IRT-theta scoring of SCCs predicting neuropsychological test performance, at selected levels of total SCC scores

	SCC total score = 1 (<i>n</i> = 265)		SCC total score = 5 (<i>n</i> = 119)	
	Standardized β	p-value	Standardized β	p-value
Neuropsychological test				
<i>Global cognition</i>				
MMSE	-0.11	.08	-0.08	.36
<i>Memory</i>				
WMS-R Logical Memory II	-0.07	.27	-0.01	.93
WMS-R Visual Reproduction II	-0.13	.04	-0.09	.36
FULD-OME	0.01	.84	-0.07	.44
<i>Executive functions</i>				
Trail Making B (s.)*	0.03	.66	0.18	.07
Clock drawing	0.06	.38	-0.19	.05
<i>Language</i>				
Animal fluency	-0.06	.35	-0.07	.48
Letter fluency	-0.06	.38	-0.17	.07
<i>Visuospatial construction</i>				
WAIS-III Block Design	-0.01	.87	0.06	.55
<i>Psychomotor speed</i>				
Trail Making A (s.)*	0.12	.05	0.16	.09

Abbreviations: IRT = Item Response Theory; SCC = Subjective Cognitive Complaints; MMSE = Mini Mental State Examination; WMS-R = Wechsler Memory Scale – Revised; FULD-OME = Fuld Object Memory Evaluation; WAIS-III = Wechsler Adult Intelligence Scale 3rd Edition.

* Higher values indicate worse performance.