Umeå University

This is an accepted version of a paper published in *Scandinavian Journal of Psychology*.

Citation for the published paper:

Vestergren, P., Rönnlund, M., Nyberg, L., Nilsson, L. (2012)

"Multigroup confirmatory factor analysis of the cognitive dysfunction questionnaire: instrument refinement and measurement invariance across age and sex"

Scandinavian Journal of Psychology, 53(5): 390-400

URL: <http://dx.doi.org/10.1111/j.1467-9450.2012.00970.x>

Access to the published version may require subscription.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-45671>

<http://umu.diva-portal.org>

Multigroup Confirmatory Factor Analysis of the Cognitive Dysfunction
Questionnaire: Instrument Refinement and Measurement Invariance across Age and
Sex

Peter Vestergren

Department of Applied Educational Science, Umeå University, Sweden

Michael Rönnlund

Department of Psychology, Umeå University, Sweden

Lars Nyberg

Department of Radiation Sciences and Integrative Medical Biology, Umeå

University, Sweden

Lars-Göran Nilsson

Department of Psychology, Stockholm University and Stockholm Brain Institute,

Sweden

Author Note

The Betula Study is funded by the Bank of Sweden Tercentenary Foundation (1988-0082:17; J2001-0682), Swedish Council for Planning and Coordination of Research (D1988-0092, D1989-0115, D1990-0074, D1991-0258, D1992-0143, D1997-0756, D1997-1841, D1999-0739, B1999-474), Swedish Council for Research in the Humanities and Social Sciences (F377/1988-2000), the Swedish Council for Social Research (1988-1990: 88-0082, and 311/1991-2000), and the Swedish Research Council (345-2003-3883, 315-2004-6977). Correspondence concerning this article should be addressed to **the current address: Peter Vestergren,**

Department of Integrative Medical Biology, Physiology section, Umeå University, S-90187 Umeå, Sweden; E-mail: peter.vestergren@physiol.umu.se

Abstract

The study adopted CFA to investigate the factorial structure and reduce the number of items of the Cognitive Dysfunction Questionnaire (CDQ; Vestergren, Rönnlund, Nyberg, & Nilsson, 2011). The analyses were based on data for a total of 1115 participants from population based samples (mean age: 63.0 ± 14.5 years, range: 25 - 95) randomly split into a refinement ($n = 569$) and a cross-validation ($n = 546$) sample. Equivalence of the measurement and structural portions of the refined model was demonstrated across the refinement and cross-validation samples. Among competing models the best fitting and parsimonious model had a hierarchical factor structure with five first-order and one second-order general factor. For the final version of the CDQ, 20 items within five domains were selected (Procedural actions, Semantic word knowledge, Face recognition, Temporal orientation, and Spatial navigation). Internal consistency reliabilities were adequate for the total scale and for the subscales. Multigroup CFAs indicated measurement invariance across age and sex up to the scalar level. Finally, higher levels of cognitive dysfunction as reflected by CDQ scores were predicted by advancing age, fewer years of education, and with deficits in general cognitive functioning as reflected by scores on the Mini-Mental State Examination. In conclusion, the CDQ appears to be psychometrically sound and shows the expected relationships with variables known to be associated with cognitive dysfunction and dementia. Future studies should apply it among clinical groups to further test its usefulness.

KEYWORDS: subjective memory, self-report measures, cognitive functioning, cognitive impairment

Multigroup Confirmatory Factor Analysis of the Cognitive Dysfunction
Questionnaire: Instrument Refinement and Measurement Invariance across Age and
Sex

The relationships between subjective (self-reported) and laboratory measures of cognitive functioning are generally weak (for reviews from different perspectives, see e.g., Gilewski & Zelinski, 1986; Herrmann, 1982; Jonker, Geerlings, & Schmaud, 2000; Perez Garcia, Godoy Garcia, Vera Guerrero, Laserna Triguero, & Puente, 1998; Reid & MacLulich, 2006). Several explanations for the weak relationships have been suggested, e.g. methodological differences between questionnaires and laboratory measures (e.g., Herrmann, 1982; Herrmann, Sheets, Gruneberg, & Torres, 2005; Rabbitt & Abson, 1990), the memory introspection paradox (e.g., Herrmann, 1982; Rabbitt & Abson, 1990), influence from social desirability, response sets, and wording of items (Gilewski & Zelinski, 1986), and individual differences in interpretation and use of self-rating scales (e.g., Cavanaugh, Feldman, & Hertzog, 1998; Rabbitt & Abson, 1990). The majority of variability in people's responses is thus not explained by cognitive test performance, whereas a common finding is that there is a relationship with depressive symptoms (e.g., Bolla, Lindgren, Bonaccorsy, & Bleecker, 1991; Zandi, 2004). A relationship with personality variables has also been observed in several studies (e.g., Hänninen, et al., 1994; Pearman & Storandt, 2004; Ponds & Jolles, 1996).

With these results in mind one could get the impression that self-report questionnaires have limited utility for assessing cognitive functioning as defined by laboratory tests. However, several studies have indicated that self-reports predict a future dementia diagnosis to some extent (e.g., Geerlings, Jonker, Bouter, Ader, & Schmand, 1999; Guarch, Marcos, Salamero, & Blesa, 2004; Schmand, Jonker,

Hooijer, & Lindeboom, 1996). Hence, it appears that self-reports, albeit suffering from methodological difficulties and being prone to influences from other variables, may be used as a source of information regarding cognitive functioning. The influence of other variables, such as depressive symptoms, may be regarded as sources of construct irrelevant variance or error that should be avoided in the development of new self-report measures.

This was the background for the development of the Cognitive Dysfunction Questionnaire (CDQ; Vestergren, et al., 2011). The intent was to create an instrument that would correlate with objective measures of memory and cognition. The rationale was that, if successful, it could have utility as a screening instrument for early signs of Alzheimer's disease (AD) in the community and clinical settings. A global construct of cognitive functioning with an emphasis on memory systems was adopted as theoretical framework for item construction. An item pool was created and clusters of items that correlated with a comprehensive set of laboratory measures of memory and cognition were selected. The clusters were derived from principal components analyses and named Procedural actions, Semantic word knowledge, Face recognition, Temporal orientation, Spatial navigation and Episodic memory for places. Early attempts at validation showed promise as the CDQ exhibited high internal consistency, and the total score correlated with age ($r = 0.40, p < .001$) and laboratory measures of memory and cognition at zero-order, and after controlling for age and years of education jointly. Furthermore, there was no significant correlation with depressive symptoms.

A new version of the CDQ was created by constructing and adding twenty new items in the six domains. By doing this it would be possible to exclude items that were similar in wording, and to provide an equal number of items for each

subscale (some components were defined by few items, and the aim was to create a short and reliable scale with a simple scoring procedure). The new version was administered in independent samples from the adult general population. The present study was concerned with refining the CDQ into a final version and investigating a number of psychometric properties. This will provide a basis for further studies on validity evidence by relations to other constructs and measures. Furthermore, it will provide valuable information regarding factorial structure, scale statistics and typical scores for the adult general population as a reference for further studies in various clinical samples. Thereby, for this study, we only included participants with no indication of cognitive impairment (i.e., Mini Mental State Examination > 24). This restricts the variability in the samples, and thereby the expected correlation between scores on the CDQ and the MMSE. At this stage, we do not address the issue of sensitivity and specificity to cognitive impairment.

The specific objectives of the study were to 1) investigate the factorial structure of the CDQ with confirmatory factor analysis, 2) perform an instrument refinement through item reduction, 3) cross-validate the factor structure of the refined version and evaluate the fit of competing models, 4) investigate measurement invariance across age and sex, and 5) to provide validity evidence by investigating relationships with age, years of education and one measure of general cognitive functioning.

Method

Participants

The participants were from the Betula prospective cohort study (Nilsson, et al., 2004; Nilsson, et al., 2000; Nilsson, et al., 1997). The Betula study was initiated in 1988, with four additional waves of data collection undertaken at

five-year intervals. The current study was based on data from samples 1 (S1), 3 (S3), and 6 (S6) at Time 5 (T5) during 2009-2010. S1 previously participated at T1-T4, and S3 at T2-T4. For S6, T5 was the first occasion for data collection. The participants were 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or 85 years of age at inclusion. S1 and S3 originally comprised 1000 participants each, and at T5 the sample sizes were reduced to 433 and 476 respectively due to attrition (death, refusal to participate further, et cetera). S6 comprised 357 participants, yielding a total of 1266 participants.

The CDQ was administered for the first time at T5, and data was obtained for 1140 of the participants. Of these, six participants (0.5%) had over 50% missing responses and were excluded from further analysis. Among the remainder, nineteen participants (1.7%) had an MMSE score < 24 , indicative of cognitive impairment (Lopez, Charter, Mostafavi, Nibut, & Smith, 2005) and were therefore also excluded. Thus, the final samples comprised 1115 participants for the analyses ($n_{S1} = 371$, $n_{S3} = 391$, $n_{S6} = 353$).

Analysis samples

To admit cross-validation the original samples were pooled and then two new samples, A ($n = 569$) and B ($n = 546$) were created by random assignment. Table 1 displays demographic and background variables for the samples. There were no significant differences between the samples on any of the variables ($p > .05$).

Instruments

The Cognitive Dysfunction Questionnaire (CDQ). The CDQ is a self-report measure of cognitive failures in everyday life. The items query about frequency of cognitive failures during the past year in six domains. There are five response alternatives: *very seldom*, *seldom*, *sometimes*, *often*, and *very often*.

Responses are assigned integers from 1 to 5 for scoring. In the development phase (Vestergren, et al., 2011), exploratory factor analyses (Principal components) suggested six positively correlated first-order factors, and thereby one second-order general factor. Internal consistency reliability was adequate for the total scale (Cronbach's $\alpha = .90$). For the present study, a total of twenty new items were constructed in the six domains: Procedural actions (5 old + 4 new items), Semantic word knowledge (4 old + 4 new items), Face recognition (3 old + 3 new items), Temporal orientation (3 old + 2 new items), Spatial navigation (3 old + 4 new items), and Episodic memory for places (2 old + 3 new items), resulting in 40 items in total.

General cognitive functioning. The Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) was used as a measure of general cognitive functioning. The MMSE is a widely used test in both clinical and research settings. It has been evaluated as a screening test for dementia in clinical settings (Tangalos, et al., 1996) and population based samples (Jacqmin-Gadda, Fabrigoule, Commenges, & Dartigues, 1997; McDowell, Kristjansson, Hill, & Hebert, 1997; Tombaugh, McDowell, Kristjansson, & Hubley, 1996). The MMSE consists of a total of 30 items assessing temporal and spatial orientation, short-term memory, episodic long-term memory, subtraction, construction, and language ability. McDowell, et al. (1997) have reported satisfactory internal consistency reliability estimates with a Cronbach's $\alpha = .78$. However, $\alpha = .40$ has been reported in a Betula sample (Rönnlund & Nilsson, 2006). A score of 24 is commonly used as a cut-off for screening purposes, i.e., a score ≤ 23 suggests cognitive impairment (Lopez, et al., 2005). Cronbach's α and other classical reliability estimates are however not fully adequate in criterion-referenced testing. Lopez, Charter, Mostafavi, Nibut, and Smith (2005) have presented alternative internal consistency reliability coefficients

(Livingstons $r = .803$ and $.795$ and Gilmer-Feldt reliability = $.764$ and $.747$) for the cut-off score of 24 on two versions of the MMSE.

Procedure

Recruitment was performed by stratified (by age and sex) random sampling through the population registry of Umeå in the north of Sweden. Prior to being included, potential participants were screened with the exclusion criteria of: dementia diagnosis, major sensory impairment, mental retardation and native tongue other than Swedish (see Nilsson et al., 1997, for additional details of recruitment, screening and measures used in the Betula study). The first step in the protocol consisted of a health examination and interviews about health status at the Betula test center. The participants were also given a collection of questionnaires, including the CDQ, to complete at home and return a week later at a follow-up cognitive testing session. Testing at the latter session lasted about 1 ½ hour and was performed by trained psychometricians.

Data Analysis

General overview of the analyses. The first step was to investigate whether the overall factor structure proposed at the development of the CDQ was plausible, and to make theory based adjustments to the model otherwise. Model fit was expected to be less than ideal as half of the items were new. Sample A was used.

The second step concerned instrument refinement, where items contributing to model misfit were excluded, and items with desirable properties were selected. Finally, the fit of the refined model was evaluated. Sample A was used.

The third step was cross-validation by means of multigroup CFA whereby the equality of the model derived from the first two steps was tested across samples A and B.

The fourth step involved testing the refined model against competing models using the pooled samples A and B. In addition to the refined model (hierarchical with five first-order factors and one second-order general factor), a single-factor model, an uncorrelated five factor model, and a correlated five factor model was tested.

In the fifth step, measurement invariance across age and sex was evaluated for the model that was judged to be the most adequate representation of the CDQ in step four. Samples A and B were pooled to one sample for this purpose.

Finally, scale statistics, differences in means between age and sex groups, and some validity evidence was investigated for the final version of the CDQ.

Statistical methods. Basic statistical analyses and imputation of missing data were conducted with PASW statistics 18.0.1. Single- and multi-group Confirmatory Factor Analysis (CFA) was performed with LISREL 8.80 (Jöreskog & Sörbom, 2006) using Robust Maximum Likelihood Estimation (RMLE) on the variance-covariance matrix. For the multigroup analyses item means were also included. RMLE was chosen because data deviated from a normal distribution, and differs from the ordinary MLE by adopting the Satorra-Bentler (S-B) scaling procedure (Chou, Bentler, & Satorra, 1991) that adjusts standard error estimates and model χ^2 based on observed non-normality of the data. Even though a non-significant p -value for χ^2 theoretically implies a well-fitting model, it is widely recognized that this is rarely achieved in applied research as it is dependent on sample size (e.g., Cheung & Rensvold, 2002). Therefore, a two-index presentation strategy was adopted for evaluation of model fit (Hu & Bentler, 1999). In addition to the χ^2 goodness-of-fit test statistic (with S-B correction) with df and p -value, the Root

Mean Squared Error of Approximation (RMSEA) with 90% confidence intervals, the Standardized Root Mean squared Residuals (SRMR) and the Comparative Fit Index (CFI) were used to evaluate model fit. Comparisons of nested models were also based the difference in χ^2 ($\Delta S-B\chi^2$) with p -value. Relatively well-fitting models have small χ^2 , nonsignificant p (usually $>.05$), RMSEA close to or below .06, SRMR close to or below .08, and CFI close to or below .95 (Hu & Bentler, 1999). The difference in $S-B\chi^2$ between nested models was computed according to the formula in Satorra and Bentler (2001).

First-order factors were scaled by fixing the regression weight of the first item to unity. The variance of the second-order factor was fixed to unity to admit model identification. For the multigroup analyses the second-order factor was scaled by fixing the regression weight of the highest loading factor to unity and thereby allowing the variance to be freely estimated for each group separately.

Tests of measurement invariance. Tests of measurement invariance or equivalence investigate whether the same construct has been measured across different groups (Chen, Sousa, & West, 2005), based on for example, age (e.g., Nyberg, et al., 2003) or sex (e.g., Maitland, Intrieri, Schaie, & Willis, 2000). The same procedures can also be adopted to investigate whether the factorial structure replicate across independent samples from the same population, i.e., cross-validation (Byrne, 2010). For these analyses, we generally adopted the nomenclature and recommendations in Vandenberg and Lance (2000) for measurement invariance in general, and Chen, Sousa and West (2005) and Dimitrov (2010) with the specific application to second-order factor models.

The aims of a specific study should determine which tests of measurement invariance are undertaken (Vandenberg & Lance, 2000). For the cross-

validation purpose we tested all recommended invariance constraints (models 0 to 6 below). For the tests of invariance across age and sex in the context of construct validation we tested models 0 to 4a below (see Dimitrov, 2010). Establishment of configural (model 1), metric (models 2a and 2b) and scalar invariance (models 3a and 3b) suggest that the same construct is measured across groups and that the units and origins of the scale are the same. This permits analyses of relations between constructs and comparisons of group means. For the second-order model tested in this study, invariant uniquenesses of first order factors (model 4a) were also tested. The test of equal error variances of items (model 4b) may be regarded as being overly stringent (Byrne, 2010). Factor variances (model 5) are not expected to be equal in different populations in general (Chen, Sousa, & West, 2005), and we expected means of the second-order factor (model 6) not to be equal.

Model 0: Invariant Covariance. The first step was an omnibus test for equality of covariance matrices. A failure to reject such a null-hypothesis warrants no further steps, i.e. complete measurement invariance is implicated (Vandenberg & Lance, 2000). If this test suggested that covariance matrices were not equal across groups, the subsequent analyses were performed through a nested hierarchy of models with increasing cross-group equality constraints on parameters (increasing degrees of freedom). Hence, each new model also contained the constraints in the previous model.

Model 1: Configural Invariance. In this step the pattern of free and fixed factor loadings were constrained to be the same across groups. This model also served as a baseline in the nested sequence. A well-fitting baseline model may be interpreted as evidence that the factor structure maps the same conceptual frame of reference used for the item responses in the groups.

Models 2a and 2b: Metric invariance. Structural relationships (first-order, and first-order and second-order factor loadings) were constrained to be equal. These models were tests of equal scaling units across groups. The presence of a second-order factor warrants testing first-order factors first.

Models 3a and 3b: Scalar invariance. The item, and item and first-order factor intercepts were constrained to be equal. As for metric invariance, the presence of a second-order factor warrants testing item intercepts first. Establishment of scalar invariance indicates that observed scores are related to latent scores in such a way that group differences in means can be meaningfully compared.

Models 4a and 4b: Invariant uniquenesses. The uniquenesses (error and unique variances) of first-order factors, and first-order factors and items were constrained to be equal. As for metric and scalar invariance, the presence of a second-order factor warrants testing either first-order factors or items first.

Model 5: Invariant factor variances. The variances of the second-order factor were constrained to be equal. The variances of first-order factors are dependent on the second-order factor and cannot be directly compared.

Model 6: Equal factor means. Means of the second-order factor was constrained to be equal. The means of first-order factors are dependent on the second-order factor mean and cannot be directly compared.

Changes in model fit as a result of equality constraints were evaluated in comparison with the previous model in the sequence. Usually, the difference in χ^2 with the associated Δdf is used to test whether a null hypothesis of invariance should be rejected ($\Delta\chi^2$ -rule). However, it has been argued that this may be a too stringent criterion as $\Delta\chi^2$, likewise to χ^2 , is dependent on sample size (e.g., Brannick, 1995; Kelloway, 1995) and therefore suggest rejection of models with trivial practical

misfit in large samples. Therefore, we also used the Δ CFI-rule as recommended by Cheung and Rensvold (2002). A difference in CFI larger than .01 is to be taken as an indication that the null hypothesis of invariance should be rejected. Along with these criteria, we also evaluated the overall fit and change in fit of the models with the alternative fit indexes. Specifically, Chen (2007) recommended a change of $\geq -.010$ in CFI, supplemented by a change of $\geq .015$ in RMSEA or a change of $\geq .030$ in SRMR as indications of non-invariance when testing loadings. For testing intercept or residual invariance, the levels are $\geq -.010$ in CFI, supplemented by a change of $\geq .015$ in RMSEA or a change of $\geq .010$ in SRMR.

Evidence of invariance should be followed by analyses aimed at identifying the source. However, as the adoption of the different criteria resulted in conflicting conclusions we performed all tests of measurement invariance.

Furthermore, it has been argued that the omnibus test often leads to contradictory findings due to the lack of a baseline model for the test of invariant variance-covariance matrices (e.g., Byrne, 2010). We therefore performed all of the subsequent tests of measurement invariance even if the hypothesis of invariant variance-covariance matrices was not rejected.

Hypothesized model. The conceptual model proposed at the development of the CDQ consisted of six domains of cognitive functioning, all part of a global construct of cognitive functioning. This implies computation of sub-scores and an overall score for the instrument. The statistical model was hierarchical and consisted of 6 first-order factors and a higher-order general factor. Each item has a nonzero loading on the first-order factor it was intended to measure, and zero loadings on all other first-order factors. The error terms for each of the items are

uncorrelated. The covariation among the first-order factors is fully explained by the second-order factor. This model is hereby denoted Model A.

One a-priori concern with Model A regarded the uniqueness of the factors *Episodic memory for places* and *Spatial navigation*. Analyses performed during the development phase suggested that they may in fact be so closely related that they should be collapsed into one *Spatial navigation*-factor. This appears plausible as navigation in a place and memory for the place likely interact in everyday life.

Instrument refinement. Items for the refined version were selected based on the criteria of a) item content, b) standardized regression weight (loading), c) absence of cross-loading, d) absence of inter-item dependency for first-order factors, and e) response distributions. Modification indexes and standardized residuals were inspected to identify suboptimal items. The modification indexes chiefly concerned suggestions of added error covariances among items in first-order factors (inter-item dependency) and between items in separate first-order factors (cross-loading).

According to Floyd and Widaman (1995) there is seldom a need for more than 20 items for a scale to achieve adequate reliability. Furthermore, there appears to be common agreement that at least three items per factor is desirable for CFA, but more is better (Marsh, Hau, Balla, & Grayson, 1998). Hence, the aim was to retain four items per factor for equal length, and as a compromise between the needs for brevity, reliability and factor identification.

Results

Distributions and missing data

The number of missing responses for individual items ranged from 0 to 1.7%. Partially missing responses were imputed by multiple regression where all other items served as predictors. The imputed values were restricted to integers in the range 1 - 5.

All items exhibited values of univariate skewness ($M = 2.0$, range: 0.1 – 7.7) and kurtosis ($M = 7.5$, range: -0.8 – 65.6) indicative of a deviation from normality. Mardia's normalized multivariate kurtosis (= 1211.3) also indicated that data deviated from multivariate normality.

Generally, positively skewed items also exhibited positive kurtosis, and several of them had a restricted range of responses. Inspection of the content of the most extremely non-normal items revealed that they described very uncommon/severe cognitive failures, for example, item 28: "You don't recall whether it is in the morning or evening" (range: 1 – 3, skewness: 7.8, and kurtosis: 66.3), and item 36: "You don't recognize a close relative" (range: 1 – 4, skewness: 5.5, and kurtosis: 33.8). These and three additional items were all new, and excluded from further analysis.

The number of responses in the alternatives *often* and *very often* were few for the majority of items, entailing several outliers. These alternatives were therefore collapsed with the *sometimes* alternative for all items in the statistical analyses. This reduced the deviations from normality: mean skewness: 1.3 (range - 0.4 - 5.0) and mean kurtosis: 2.1 (range -1.5 – 26.8), and Mardia's normalized multivariate kurtosis (= 426.1).

Confirmatory factor analysis and refinement – sample A

Generally, the analyses were complicated by a small determinant of the covariance matrix. This was likely a result of rather highly correlated items and

thereby possible multicollinearity. However, no subsets of specifically problematic items, or combinations of items were identified. Rather, it was judged that the overall similarity of the phrasings in the items was the cause.

The first run of Model A resulted in an error variance/uniqueness estimate not significantly different from zero for the factor *Episodic memory for places*. The standardized regression weight from the general factor to this factor was close to unity (0.96). Inspection of covariance estimates for the “Y-model” in LISREL revealed that the highest correlation was between the factors *Episodic memory for places* and *Spatial navigation* (0.84). Given these results, and the results from the previous study (Vestergren, et al., 2011), it was concluded that the two factors should be collapsed into one factor. Consequently, the *Episodic memory for places* factor was excluded and the items were assigned as indicators of the Spatial navigation factor.

In the second run, the respecified five factor model (B) resulted in parameter estimates that were all significant ($p < .05$). The model was bordering on good fit as judged by alternative fit-indexes and therefore deemed adequate given that half of the items were new ($S-B\chi^2 = 1980.9$, $df = 555$, $p < .000$, RMSEA = .067 (CI : .064 - .070), SRMR = .068, CFI = .96). R^2 was .78 for Procedural actions, .66 for Semantic word knowledge, .35 for Face recognition, .37 for Temporal orientation, and .80 for Spatial navigation. Inspection of modification indexes and residuals suggested that improved fit could be achieved mainly by allowing non-zero error covariances between items within first-order factors. Hence, there was systematic variance shared by items and not accounted for by the first-order factors. Each such item-pair was inspected with regard to content, standardized regression weight and distribution. In the majority of cases similar wording was clearly the

cause of the misfit. Subsequently, given redundancy in content the item with the highest value of skewness and kurtosis and/or lowest standardized regression weight was excluded. Four items per subscale were selected, resulting in twenty items in total.

The final, refined model (C) had good fit ($S-B\chi^2 = 357.6$, $df=165$, $p < .000$, RMSEA=.045 (CI ; .039 - .052), SRMR =.061, CFI = .98) and all parameter estimates were significant. Figure 1 displays the path diagram with standardized parameter estimates (regression weights and unique / error variances) for Model C in sample A. R^2 was .79 for Procedural actions, .66 for Semantic word knowledge, .33 for Face recognition, .37 for Temporal orientation, and .66 for Spatial navigation.

Crossvalidation (Samples A and B)

Model C was cross-validated by means of multigroup CFA using Samples A and B. As described previously the fit of the model was good in Sample A, and the results suggested comparable fit in Sample B ($S-B\chi^2 = 374.0$, $df=165$, $p < .000$, RMSEA=.048 (CI ; .042 - .055), SRMR =.062, CFI = .98). The omnibus test suggested equality of covariance matrices as reflected in excellent fit ($S-B\chi^2 = 230.5$, $df=210$, $p > .05$, RMSEA=.013 (CI ; .000 - .023), SRMR =.045 and .047, CFI = 1.00). None of the equality constraints resulted in a significant worsening of fit as judged by either $\Delta S-B\chi^2$ or ΔCFI , and the alternative fit indexes suggested well-fitting models. Given the overall results it was concluded that Model C was well-fitting and operating equivalently across the samples. The subsequent analyses were based on the pooled samples.

Competing models (Samples A+B)

It is warranted to test the fit of alternative representations of the factorial structure for a measurement instrument, i.e., competing models. There were

no obvious alternative models suggested by theory as the domains were mainly selected based on correlations with criterion measures in the first phase of development (Vestergren, et al., 2011). The competing models tested were therefore variations differing in parsimony; a correlated five-factor model (D), an uncorrelated five-factor model (E) and a single-factor model (F). Models C, E and F are (parameter) nested in model D as restrictions on parameters in model D can result in the other models (see e.g., Bentler & Satorra, 2008). Models E and F have the same *df* and are both nested in models D and C, but not within each other. Neither are they equivalent.

The results from the analyses of competing models are summarized in Table 2. The correlated five-factor model provided a good fit to the data. The nested hierarchical model had 5 additional *df*, and a significant $\Delta S-B\chi^2$ suggested that the fit was worse. However, $\Delta CFI < .01$ and very modest differences in alternative fit indexes indicated that the hierarchical model was equally well-fitting. The single-factor and uncorrelated five-factor models were both bad-fitting according to all criteria. Even though the correlated five-factor model had slightly better fit, based on parsimony (less parameters) the hierarchical model (C) was considered the best for describing the structure of the CDQ.

Measurement invariance (sample A+B)

Multi-group CFA was performed to investigate whether the measurement and structural portions of Model C were invariant over age and sex. In addition to the grouping based on sex: female ($n = 599$) and male ($n = 516$), a grouping was also made according to age, 25 - 60 years ($n = 501$) and 65 - 95 years ($n = 614$)¹.

Age invariance. Table 3 displays the results from the analyses of age-invariance. Model C had adequate fit in both age-groups when fitted separately. However, the omnibus test suggested rejection of the null-hypothesis of equality of covariance matrices. Model fit was good judging by values of RMSEA and CFI, but not $S-B\chi^2$ and SRMR. Configural invariance was supported by a well-fitting baseline model, and none of the subsequent null-hypotheses of measurement invariance were rejected using the ΔCFI -rule. In contrast, model 2b (metric invariance of first- and second-order factor loadings) and 3a (scalar invariance for intercepts of items) were rejected using the $\Delta S-B\chi^2$ -rule.

To further investigate the possible sources of invariance as suggested by the $\Delta S-B\chi^2$ -rule, modification indexes were investigated for the models with statistically significant worsening of fit. Parameters were then freed until $\Delta S-B\chi^2$ was no longer statistically significant ($p > .05$). For Model 2b, this meant freeing the path from the second-order general factor to the Temporal orientation factor. The standardized estimate was higher in the old group (.81) compared to the young group (.44).

For model 3a, it meant freeing the intercepts of items 11 - “don’t recall in what order to do something you’ve done many times before” (old group: 1.33 vs. young group: 1.42), 22 - “take the wrong route when you’re going someplace” (old group: 1.38 vs. young group: 1.47) and 23 - “don’t recognize a person you’ve met many times before” (old group: 1.31 vs. young group: 1.19).

Sex invariance. Table 4 displays the results from the analysis of sex-invariance. Model C had adequate fit in both sex-groups when fitted separately. The omnibus test suggested rejection of the null-hypothesis of equality of covariance matrices. Model fit was good judging by alternative fit indexes, but $S-B\chi^2$ was

statistically significant. The baseline model was well-fitting and thereby supported configural invariance. As for the results from the analysis of age-invariance, none of the subsequent null-hypotheses of measurement invariance were rejected using the Δ CFI-rule. Models 2a (metric invariance for first-order factor loadings), 3a (scalar invariance of intercepts of items), and 3b (scalar invariance of intercepts of items and first-order factors) were rejected using the Δ S-B χ^2 -rule.

For Model 2a, freeing the loading of item 14 - “don’t recall that you’ve been to a place before” on the spatial navigation factor was needed to achieve a non-significant increase in S-B χ^2 . The standardized estimate was higher in the male group (.89) than in the female group (.57).

For model 3a, the intercepts of three items were freed: 6 - “get lost” (Female: 2.35 vs. Male: 2.17); 27 - “don’t recall the use of a word” (Female: 2.18 vs. Male: 2.34); and 31 - “don’t recall the spelling of a word” (Female: 2.05 vs. Male: 2.26).

For model 3b, freeing the intercepts of the spatial navigation factor was needed to attain a non significant increase in S-B χ^2 (female: 0 and male: -0.04).

Item and scale statistics

Appendix A contains the final Swedish version of the CDQ. The items have been renumbered. Appendix B contains the corresponding version freely translated to English. Table 5 displays item statistics for the final version of the CDQ. Cronbach’s α was .89 for the total scale, and for respective subscale: Procedural actions, .80; Semantic word knowledge, .73; Face recognition, .86; Temporal orientation, .67; and Spatial navigation, .76. Coefficient α is however inappropriate as a single summary of the internal consistency of a composite score (Bentler, 2009). Therefore, the alternative structural equation model-based Raykov's

reliability rho (Raykov, 1997) was also computed, resulting in slightly higher values, .90 for the total scale, and for respective subscale: Procedural actions, .81; Semantic word knowledge, .74; Face recognition, .88; Temporal orientation, .69; and Spatial navigation, .76. Theoretically, the possible range of scores on the CDQ is 20 – 100. The observed range was 20 – 75. The scale mean was 29.0 ± 7.8 .

Differences in means between age and sex groups

A 2 X 2 ANOVA with age-group (25 – 60 and 65 – 95 years of age) and sex as between group factors with CDQ score as dependent variable was performed. There were significant main effects of age, $F(1,1115) = 34.38, p < .001, \eta^2 = .030$, and sex, $F(1,1115) = 5.0, p < .05, \eta^2 = .005$. There was also a significant interaction-effect $F(1,1115) = 5.5, p < .05, \eta^2 = .005$, due to old males having higher CDQ-scores compared to old females ($M: 31.2, CI: 30.3 - 32.1$ vs. $M: 29.0, CI: 28.2 - 29.9$), whereas the difference was modest for young males and females ($M: 27.4, CI: 26.5 - 28.4$ vs. $M: 27.4, CI: 26.4 - 28.3$).

Validity evidence

Validity evidence was investigated with multiple regression analysis with CDQ score as dependent variable and MMSE-score, age, and years of education as independent variables. All variables were significant predictors in the model ($F(3, 1102) = 19.6, p < .001, R^2 = .051$). Higher scores on the CDQ was associated with lower scores on the MMSE ($\beta = -.11, t = -3.7, p < .001$), higher age ($\beta = .12, t = 3.5, p < .001$), and fewer years of education ($\beta = -.07, t = -2.1, p < .05$).

Discussion

The objectives of this study were to investigate the factorial structure of the CDQ, to perform an instrument refinement, to cross-validate the factor structure, to evaluate the fit of competing models, to investigate measurement invariance

across age and sex, and to provide validity evidence by investigating relationships with age, years of education and one measure of general cognitive functioning.

Beginning with the factorial structure, the hypothesized initial model with six first-order factors and a single higher-order factor was not supported by the data. The major misspecification concerned the factors Episodic memory for places and Spatial navigation that were subsequently collapsed into one Spatial navigation factor. We consider this modification non-controversial in light of similar content. The resulting hierarchical model with five first-order factors and one second-order factor was equivalent and well-fitting across the refinement and cross-validation samples. The results from the analysis of competing models showed that the correlated five factor model had slightly better fit than the hierarchical model. However, the difference in fit was modest and the hierarchical model is more parsimonious. Furthermore, this factor structure is beneficial as it maps computations of sub-scores (corresponding to first-order factors) and a total score (corresponding to a general factor) for the scale. The hierarchical model was therefore deemed the best representation of the factorial structure of the CDQ.

Establishment of configural, metric and scalar invariance suggest that the same construct is measured across groups and that the units and origins of the scale are the same. This permits analyses of relations between constructs and comparisons of group means. The results show that measurement invariance across age and sex is plausible and that our subsequent analyses of differences in mean scores across age and sex groups are justified. At the same time, there is some ambiguity in the results when considering that the $\Delta S-B\chi^2$ -rule suggested that few of the equality constraints were tenable, but the ΔCFI -rule and alternative fit indexes suggested well-fitting models and minimal changes in fit as a result of equality

constraints. The ambiguity is likely because the ΔCFI -rule (Cheung & Rensvold, 2002) may be a liberal test with insufficient power to detect invariance, whereas the $\Delta\chi^2$ -rule may be too conservative (Chen, et al., 2005). It is difficult to judge where the ideal or moderate zone is, but our final judgment is that measurement invariance across age and sex up to the scalar level is plausible. This conclusion is supported by the subsequent analyses aimed at finding the sources of possible invariance; there were a few parameters that were responsible for significant increases in $\text{S-B}\chi^2$.

A few caveats, or concerns, regarding the construction and refinement of the CDQ may be pointed out. In regard to the composition of the instrument, it could be of potential concern that the items within the domains are rather similar in content. One could possibly argue that more dissimilar items should have been included in the instrument, or that the reduction of the number of items should have been even more extensive. However, the results from the analyses during the first step of development (Vestergren, et al., 2011), suggested that the domains that were most related to objective test performance on cognitive tests were narrow and specific in content. As in any instrument development process, the final subset of items is the result of compromises between theory, semantic content, and statistical properties.

Also, it should be noted that the instrument has a floor effect, and that this is more pronounced in younger individuals (i.e. under 65 years of age). A difference in endorsement of cognitive failures is expected and desirable across age. However, the rather high numbers of individuals that endorse very few cognitive failures result in non-normal item distributions. This study was however based on a healthy sample from the adult general population where the majority is not expected to suffer from

cognitive dysfunction. We expect higher endorsement of cognitive failures in various clinical samples.

Taken together, the results have provided validity evidence regarding the factorial structure and measurement properties of the CDQ. Evidence of construct validity based on relations to other variables was restricted to investigating those with MMSE scores, age, and years of education. Even though explained variance was modest, MMSE scores, age, and years of education predicted scores on the CDQ in the expected directions. As noted in the introduction, the inclusion criterion of $\text{MMSE} > 24$, restricts the variability in the samples, and thereby the expected correlation between scores the CDQ and MMSE. In clinical samples, with greater variability in cognitive measures, such as the MMSE, the chances of finding a more substantial relation should be higher. As in many other studies with self-report measures of memory/cognition, the majority of the variability in CDQ-scores could not be explained by objective test performance. However, the extensive analyses regarding internal structure and measurement properties in this study did not permit an equally extensive focus on relations to a fuller set of objective cognitive tests. For the same reasons, we did not investigate whether the previous finding of a lack of a relationship with depressive symptoms could be replicated for the final version of the CDQ. Forthcoming studies should address these issues thoroughly, preferably by adopting latent variable modelling.

A related issue, that is yet to be addressed, is the criterion-referenced measurement properties of the questionnaire. If the CDQ is to be used as a screening instrument for cognitive dysfunction, other aspects need to be investigated. One such issue is comparable prediction of a criterion (e.g., episodic memory impairment). From this perspective, we would also be concerned with investigating whether

prediction is of equal magnitude and accuracy across age and sex, but also across other groups (e.g., clinical).

In conclusion, the refined version of the CDQ appears to be psychometrically sound and shows the expected relationships with variables known to be associated with cognitive dysfunction and dementia. Future studies should investigate further evidence of validity by including additional measures of cognitive functioning and apply it in examinations of clinical groups to further test its usefulness.

References

- Bentler, P. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, 74(1), 137-143. doi: 10.1007/s11336-008-9100-1
- Bentler, P., & Satorra, A. (2008). Testing model nesting and equivalence. *Department of Statistics Papers*. Los Angeles: University of California.
- Bolla, K. I., Lindgren, K. N., Bonaccorsy, C., & Bleecker, M. L. (1991). Memory complaints in older adults: Fact or fiction? *Archives of Neurology*, 48(1), 61-64. doi: 10.1001/archneur.1991.00530130069022
- Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, 16(3), 201-213. doi: 10.1002/job.4030160303
- Bryant, F. B., & Satorra, A. (in press). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling: A Multidisciplinary Journal*.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS. Basic concepts, applications, and programming* (2nd ed.). New York: Routledge.
- Cavanaugh, J. C., Feldman, J. M., & Hertzog, C. (1998). Memory beliefs as social cognition: A reconceptualization of what memory questionnaires assess. *Review of General Psychology*, 2(1), 48-65. doi: 10.1037/1089-2680.2.1.48
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464-504. doi: 10.1080/10705510701301834
- Chen, F. F. Sousa, K., & West, S. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling: A*

Multidisciplinary Journal, 12(3), 471-492. doi:

10.1207/s15328007sem1203_7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for

testing measurement invariance. *Structural Equation Modeling: A*

Multidisciplinary Journal, 9(2), 233 - 255. doi:

10.1207/S15328007SEM0902_5

Chou, C., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust

standard errors for non-normal data in covariance structure analysis: A Monte

Carlo study. *British Journal of Mathematical and Statistical Psychology*, 44,

347-357.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct

validation. *Measurement and Evaluation in Counseling and Development*,

43(2), 121-149. doi: 10.1177/0748175610373459

Floyd, F., & Widaman, K. (1995). Factor analysis in the development and refinement

of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A

practical method for grading the cognitive state of patients for the clinician.

Journal of Psychiatric Research, 12(3), 189-198. doi: 10.1016/0022-

3956(75)90026-6

Geerlings, M. I., Jonker, C., Bouter, L. M., Ader, H. J., & Schmand, B. (1999).

Association between memory complaints and incident Alzheimer's disease in

elderly people with normal baseline cognition. *American Journal of*

Psychiatry, 156(4), 531-537.

- Gilewski, M. J., & Zelinski, E. M. (1986). Questionnaire assessment of memory complaints. In L. W. Poon (Ed.), *Handbook for clinical memory assessment of older adults*. Washington: American Psychological Association.
- Guarch, J., Marcos, T., Salamero, M., & Blesa, R. (2004). Neuropsychological markers of dementia in patients with memory complaints. *International Journal of Geriatric Psychiatry*, 19(4), 352-358. doi: 10.1002/gps.1074
- Herrmann, D., J. (1982). Know thy memory: The use of questionnaires to assess and study memory. *Psychological Bulletin*, 92(2), 434-452. doi: 10.1037/0033-2909.92.2.434
- Herrmann, D. J., Sheets, V., Gruneberg, M., & Torres, R. (2005). Are self reports of memory failure accurate? *Applied Cognitive Psychology*, 19(7), 821-841. doi: 10.1002/acp.1125
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1 - 55. doi: 10.1080/10705519909540118
- Hänninen, T., Reinikainen, K. J., Helkala, E. L., Koivisto, K., Mykkänen, L., Laakso, M., et al. (1994). Subjective memory complaints and personality traits in normal elderly subjects. *Journal of the American Geriatrics Society*, 42(1), 1-4.
- Jacqmin-Gadda, H., Fabrigoule, C., Commenges, D., & Dartigues, J. F. (1997). A 5-year longitudinal study of the Mini-Mental State Examination in normal aging. *American Journal of Epidemiology*, 145(6), 498-506.
- Jonker, C., Geerlings, M. I., & Schmaud, B. (2000). Are memory complaints predictive for dementia? A review of clinical and population-based studies.

- International Journal of Geriatric Psychiatry*, 15(11), 983-991. doi: 10.1002/1099-1166(200011)15:11<983::AID-GPS238>3.0.CO;2-5
- Jöreskog, K. G., & Sörbom, D. (2006). LISREL for Windows. Lincolnwood, IL: Scientific Software International, Inc.
- Kelloway, E. K. (1995). Structural equation modelling in perspective. *Journal of Organizational Behavior*, 16(3), 215-224. doi: 10.1002/job.4030160304
- Lopez, M. N., Charter, R. A., Mostafavi, B., Nibut, L. P., & Smith, W. E. (2005). Psychometric properties of the Folstein Mini-Mental State Examination. *Assessment*, 12(2), 137-144. doi: 12/2/137 [pii] 10.1177/1073191105275412
- Maitland, S. B., Intrieri, R. C., Schaie, W. K., & Willis, S. L. (2000). Gender differences and changes in cognitive abilities across the adult life span. *Aging, Neuropsychology, and Cognition*, 7(1), 32 - 53. doi: 10.1076/anec.7.1.32.807
- Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33(2), 181-220. doi: 10.1207/s15327906mbr3302_1
- McDowell, I., Kristjansson, B., Hill, G. B., & Hebert, R. (1997). Community screening for dementia: The Mini Mental State Exam (MMSE) and Modified Mini-Mental State Exam (3MS) compared. *Journal of Clinical Epidemiology*, 50(4), 377-383. doi: 10.1016/S0895-4356(97)00060-7
- Nilsson, L.-G., Adolfsson, R., Bäckman, L., de Frias, C. M., Molander, B., & Nyberg, L. (2004). Betula: A prospective cohort study on memory, health and aging. *Aging, Neuropsychology, and Cognition*, 11(2-3), 134-148. doi: 10.1080/13825580490511026

- Nilsson, L.-G., Adolfsson, R., Bäckman, L., Molander, B., Norin, E., Nyberg, L., et al. (2000). The Betula Study: Objectives, assessment and design. *International Journal of Psychology*, 35(3-4), 234-234.
- Nilsson, L.-G., Bäckman, L., Erngrund, K., Nyberg, L., Adolfsson, R., Bucht, G., et al. (1997). The Betula prospective cohort study: Memory, health and aging. *Aging, Neuropsychology, and Cognition*, 4(1), 1-32. doi: 10.1080/13825589708256633
- Nyberg, L., Maitland, S. B., Rönnlund, M., Bäckman, L., Dixon, R. A., Wahlin, Å., et al. (2003). Selective adult age differences in an age-invariant multifactor model of declarative memory. *Psychology and Aging*, 18(1), 149-160. doi: 10.1037/0882-7974.18.1.149
- Pearman, A., & Storandt, M. (2004). Predictors of subjective memory in older adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 59(1), 4-6. doi: 10.1093/geronb/59.1.P4
- Perez Garcia, M., Godoy Garcia, J. F., Vera Guerrero, N., Laserna Triguero, J. A., & Puente, A. E. (1998). Neuropsychological evaluation of everyday memory. *Neuropsychology Review*, 8(4), 203-227. doi: 10.1023/A:1021622319851
- Ponds, R. W. H. M., & Jolles, J. (1996). Memory complaints in elderly people: the role of memory abilities, metamemory, depression, and personality. *Educational Gerontology*, 22(4), 341 - 357. doi: 10.1080/0360127960220404
- Rabbitt, P., & Abson, V. (1990). "Lost and found": Some logical and methodological limitations of self-report questionnaires as tools to study cognitive ageing. *British Journal of Psychology*, 81(1), 1-16. doi: 10.1111/j.2044-8295.1990.tb02342.x

Raykov, T. (1997). Estimation of composite reliability for congeneric measures.

Applied Psychological Measurement, 21(2), 173-184. doi:

10.1177/01466216970212006

Reid, L. M., & MacLulich, A. M. J. (2006). Subjective memory complaints and cognitive impairment in older people. *Dementia and Geriatric Cognitive*

Disorders, 22(5-6), 471-485. doi: 10.1159/000096295

Rönnlund, M., & Nilsson, L.-G. (2006). The Betula study: Reliabilities and long-term stabilities. *Baltic Journal of Psychology*, 7(2 - Special Edition of VII International Baltic Psychology Conference Addresses and Presentations), 6-14.

Satorra, A., & Bentler, P. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514. doi:

10.1007/bf02296192

Schmand, B., Jonker, C., Hooijer, C., & Lindeboom, J. (1996). Subjective memory complaints may announce dementia. *Neurology*, 46(1), 121-125.

Tangalos, E. G., Smith, G. E., Ivnik, R. J., Petersen, R. C., Kokmen, E., Kurland, L. T., et al. (1996). The Mini-Mental State Examination in general medical practice: Clinical utility and acceptance. *Mayo Clinic Proceedings*, 71(9), 829-837. doi: 10.4065/71.9.829

Tombaugh, T. N., McDowell, I., Kristjansson, B., & Hubley, A. M. (1996). Mini-Mental State Examination (MMSE) and the modified MMSE (3MS): A psychometric comparison and normative data. *Psychological Assessment*, 8(1), 48-59. doi: 10.1037/1040-3590.8.1.48

Vandenberg, R. J., & Lance, C., E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and

recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70. doi: 10.1177/109442810031002

Vestergren, P., Rönnlund, M., Nyberg, L., & Nilsson, L.-G. (2011). Development of the Cognitive Dysfunction Questionnaire (CDQ) in a population based sample. *Scandinavian Journal of Psychology*, 52(3), 218-228. doi: 10.1111/j.1467-9450.2010.00861.x

Zandi, T. (2004). Relationship between subjective memory complaints, objective memory performance, and depression among older adults. *American Journal of Alzheimer's Disease and Other Dementias*, 19(6), 353-360. doi: 10.1177/153331750401900610

Footnote

¹ Due to few participants in the youngest and oldest age-cohorts the data only permitted the forming of two groups with roughly equal sample size.

Figure 1. Path diagram with standardized parameter estimates for the final refined version of the CDQ in Sample A. Robust Maximum Likelihood Estimation. (proc = Procedural actions, sem = Semantic word knowledge, face = Face recognition, temp = Temporal orientation, spat = Spatial navigation).

Table 1

Demographic and Background Variables for the Refinement (A) and Crossvalidation (B) Samples

	Sample		
	A	B	Total
	(<i>n</i> = 569)	(<i>n</i> = 546)	(<i>N</i> = 1115)
Age	63.1 (15.0)	62.9 (14.0)	63.0 (14.5)
Education (years)	12.5 (4.4)	12.3 (4.1)	12.4 (4.2)
MMSE	28.0 (1.6)	28.0 (1.6)	28.0 (1.6)
Sex (% male)	44.8	47.8	46.3

Note. *SD* within parentheses. No significant differences between samples A and B ($p > .05$)

Table 2

Results from the Analysis of Competing Models in Sample A+B (N=1115)

Model	S-B χ^2	df	RMSEA	90% CI	SRMR	CFI	Δ CFI	Δ df	Δ S-B χ^2	Model comparison
1. Correlated five factor (D)	532.9***	160	.046	.042 - .050	.054	.98				
2. Hierarchical (C)	558.5***	165	.046	.042 - .050	.057	.98	.00	5	23.4***	2 vs. 1
3. Uncorrelated five factor (E)	2232.6***	170	.104	.101 - .108	.249	.91	-.07	5	676.9***	3 vs. 2
4. One factor (F)	2290.8***	170	.106	.102 - .110	.083	.90	-.01	5	528.1***	4 vs. 2

Note. S-B χ^2 = Satorra-Bentler scaled statistic, *df* = degrees of freedom, RMSEA = Root Mean Squared Error of Approximation, 90% CI = 90% Confidence Interval for RMSEA, SRMR = Standardized Root Mean Residual, CFI = Comparative Fit Index, Δ CFI = Difference in CFI between nested models, Δ df = Difference in degrees of freedom between nested models, Δ S-B χ^2 = Difference in S-B χ^2 between nested models,

*** $p < .001$.

Table 3

Results from the Analysis of Measurement Invariance across Age-groups.

Model	S-B χ^2	df	RMSEA	90% CI	SRMR ^a	CFI	Δ CFI	Δ df	Δ S-B χ^2
C in age group 25 – 60 ^b	286.2***	165	.038	.031 - .046	.059	.984			
C in age group 65 – 95 ^c	422.6***	165	.050	.045 - .056	.063	.981			
					25 - 60	65 - 95			
0. Invariant covariance matrices	404.2***	210	.041	.035 - .047	.111	.091	.991		
1. Configural invariance (baseline model)	702.5***	330	.045	.040 - .050	.059	.066	.982		
2a. Metric invariance (first-order factor loadings)	723.4***	345	.044	.040 - .049	.065	.073	.982	.000	15
2b. Metric invariance (first- and second-order factor loadings)	761.8***	349	.046	.042 - .051	.073	.079	.980	-.002	4
3a. Scalar invariance (intercepts of items)	810.0***	364	.047	.043 - .051	.074	.077	.979	-.001	15
3b. Scalar invariance (intercepts of items and first-order factors)	828.3***	368	.047	.043 - .052	.073	.076	.978	-.001	4
4a. Invariant uniquenesses (error variances of first-order factors)	847.4***	373	.048	.044 - .052	.077	.077	.977	-.001	5

Note. S-B χ^2 = Satorra-Bentler scaled statistic, df = degrees of freedom, SRMR = Standardized Root Mean Residual, RMSEA = Root Mean Squared Error of Approximation, 90% CI = 90% Confidence Interval for RMSEA, CFI = Comparative Fit Index, Δ df = Difference in degrees of freedom between nested models, Δ S-B χ^2 = Difference in S-B χ^2 between nested models.

^aLISREL reports separate values of SRMR in multi-group analyses.

^bn=501, ^cn=614.

***p < .001.

Table 4

Results from the Analysis of Measurement Invariance across Sex

Model	S-B χ^2	df	RMSEA	90% CI	SRMR ^a	CFI	Δ CFI	Δ df	Δ S-B χ^2
C in Female group ^b	398.8***	165	.049	.043 - .055	.066	.978			
C in Male group ^c	346.4***	165	.046	.039 - .053	.060	.984			
					Female	Male			
0. Invariant covariance matrices	301.5**	210	.028	.021 - .035	.059	.068	.996		
1. Configural invariance (baseline model)	746.2***	330	.048	.043 - .052	.066	.060	.981		
2a. Metric invariance (first-order factor loadings)	788.4***	345	.048	.044 - .053	.062	.073	.980	-.001	15 29.0*
2b. Metric invariance (first- and second-order factor loadings)	800.3***	349	.048	.044 - .052	.066	.074	.979	-.001	4 7.7
3a. Scalar invariance (intercepts of items)	842.3***	364	.049	.044 - .053	.066	.074	.978	-.001	15 52.1***
3b. Scalar invariance (intercepts of items and first-order factors)	886.1***	368	.050	.046 - .055	.067	.074	.976	-.002	4 79.7***
4a. Invariant uniquenesses (error variances of first-order factors)	895.4***	373	.050	.046 - .054	.069	.075	.976	.000	5 9.0

Note. S-B χ^2 = Satorra-Bentler scaled statistic, df = degrees of freedom, SRMR = Standardized Root Mean Residual, RMSEA = Root Mean Squared Error of Approximation, 90% CI = 90% Confidence Interval for RMSEA, CFI = Comparative Fit Index, Δ df = Difference in degrees of freedom between nested models, Δ S-B χ^2 = Difference in S-B χ^2 between nested models.

^aLISREL reports separate values of SRMR in multi-group analyses.

^bn = 599, ^cn = 516.

***p < .001, **p < .01, *p < .05.

Table 5

Item Statistics for the Final Version of the CDQ

Item	Content (translated)	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	Item-Total Correlation ^a
<i>Procedural actions</i>								
11	don't recall in what order to do something you've done many times before?	1	5	1.46	0.68	1.32	1.30	.61
16	do things by old habit and discover that you fail?	1	5	1.51	0.70	1.21	0.93	.60
19	do something by routine – but discover that it is something you've already done?	1	5	1.42	0.65	1.35	1.17	.57
37	miss to perform an important step in a routine chore?	1	5	1.28	0.53	1.93	4.13	.57
<i>Semantic word knowledge</i>								
1	don't recall the meaning of a word?	1	5	1.73	0.86	0.74	-0.65	.52
27	don't recall how to use a word?	1	5	1.33	0.61	1.91	3.82	.56
31	don't recall the spelling of a word?	1	5	1.70	0.84	0.92	0.02	.38
39	use a word in the wrong way?	1	5	1.52	0.70	1.11	0.47	.55
<i>Face recognition</i>								
2	don't recognize a person you know you've met before?	1	5	1.93	0.90	0.57	-0.50	.53
8	don't recognize a face?	1	5	1.81	0.89	0.78	-0.16	.57
17	don't recall a face?	1	5	1.84	0.88	0.74	-0.15	.62
23	don't recognize a person you've met many times before?	1	5	1.33	0.64	2.13	4.97	.58
<i>Temporal orientation</i>								
3	don't recall what day of the week it is?	1	5	1.35	0.68	1.98	3.56	.47
18	don't recall what month it is?	1	5	1.11	0.39	4.20	21.14	.52
24	don't recall what year it is?	1	3	1.06	0.26	5.00	26.79	.34
33	don't recall what time of the day it is?	1	5	1.09	0.34	4.62	27.61	.44
<i>Spatial navigation</i>								
6	get lost?	1	5	1.30	0.63	2.33	5.77	.42
12	can't find your way in a place you know well?	1	5	1.21	0.51	2.62	7.55	.58
22	take the wrong route when you're going someplace?	1	4	1.48	0.71	1.28	0.72	.50
14	don't recall that you've been to a place before?	1	4	1.36	0.61	1.57	1.77	.63

Note. ^acorrected.

Appendix A. Final version of the CDQ (Swedish). Copyright 2011 by The Authors.

CDQ

Här följer ett antal frågor om händelser i vardagen. Sätt **ett** kryss per rad i rutan under det alternativ du tycker **stämmer bäst överens** med din situation under **det senaste året**. Det är viktigt att du besvarar varje fråga.

	<i>Händer det att du...</i>	<i>mycket sällan</i>	<i>sällan</i>	<i>ibland</i>	<i>ofta</i>	<i>mycket ofta</i>
1	inte kommer ihåg vad ett ord betyder?	☐	☐	☐	☐	☐
2	inte känner igen en person som du vet att du träffat förut?	☐	☐	☐	☐	☐
3	inte kommer ihåg vilken veckodag det är?	☐	☐	☐	☐	☐
4	går vilse?	☐	☐	☐	☐	☐
5	inte kommer ihåg i vilken ordning man ska göra något som du gjort många gånger förut?	☐	☐	☐	☐	☐
6	inte känner igen ett ansikte?	☐	☐	☐	☐	☐
7	inte hittar vägen på en plats du känner väl?	☐	☐	☐	☐	☐
8	gör saker av gammal vana och upptäcker att det blir fel?	☐	☐	☐	☐	☐
9	inte kommer ihåg vilken månad det är?	☐	☐	☐	☐	☐
10	inte kommer ihåg hur man brukar använda ett ord?	☐	☐	☐	☐	☐
	<i>Händer det att du...</i>	<i>mycket sällan</i>	<i>sällan</i>	<i>ibland</i>	<i>ofta</i>	<i>mycket ofta</i>
11	inte kommer ihåg att du varit på en plats förut?	☐	☐	☐	☐	☐
12	inte kommer ihåg ett ansikte?	☐	☐	☐	☐	☐
13	gör en sak på rutin - men upptäcker att det är något du redan gjort?	☐	☐	☐	☐	☐
14	inte kommer ihåg vilket år det är?	☐	☐	☐	☐	☐
15	inte kommer ihåg hur ett ord stavas?	☐	☐	☐	☐	☐
16	går fel väg när du ska någonstans?	☐	☐	☐	☐	☐
17	inte känner igen en person som du träffat många gånger förut?	☐	☐	☐	☐	☐
18	inte kommer ihåg vilken tid på dagen det är?	☐	☐	☐	☐	☐
19	missar att utföra ett viktigt steg i en rutinsyssa?	☐	☐	☐	☐	☐
20	använder ett ord på fel sätt?	☐	☐	☐	☐	☐

Appendix B. Final version of the CDQ (free translation to English). Copyright 2011 by The Authors.

CDQ

Here are a number of questions about events in everyday life. Put **one** x in each row in the square below the alternative you think **corresponds best** with your situation during the **last year**. It is important that you answer all of the questions.

	<i>Does it happen that you...</i>	<i>very seldom</i>	<i>seldom</i>	<i>sometimes</i>	<i>often</i>	<i>very often</i>
1	don't recall the meaning of a word?	☐	☐	☐	☐	☐
2	don't recognize a person you know you've met before?	☐	☐	☐	☐	☐
3	don't recall what day of the week it is?	☐	☐	☐	☐	☐
4	get lost?	☐	☐	☐	☐	☐
5	don't recall in what order to do something you've done many times before	☐	☐	☐	☐	☐
6	don't recognize a face?	☐	☐	☐	☐	☐
7	can't find your way in a place you know well?	☐	☐	☐	☐	☐
8	do things by old habit and discover that you fail?	☐	☐	☐	☐	☐
9	don't recall what month it is?	☐	☐	☐	☐	☐
1	don't recall how to use a word?	☐	☐	☐	☐	☐
	<i>Does it happen that you...</i>	<i>very seldom</i>	<i>seldom</i>	<i>sometimes</i>	<i>often</i>	<i>very often</i>
1	don't recall that you've been to a place before?	☐	☐	☐	☐	☐
1	don't recall a face?	☐	☐	☐	☐	☐
1	do something by routine – but discover that it is something you've already done?	☐	☐	☐	☐	☐
1	don't recall what year it is?	☐	☐	☐	☐	☐
1	don't recall the spelling of a word?	☐	☐	☐	☐	☐
1	take the wrong route when you're going someplace?	☐	☐	☐	☐	☐
1	don't recognize a person you've met many times before?	☐	☐	☐	☐	☐
1	don't recall what time of the day it is?	☐	☐	☐	☐	☐
1	miss to perform an important step in a routine chore?	☐	☐	☐	☐	☐
2	use a word in the wrong way?	☐	☐	☐	☐	☐