Published in final edited form as:

Int J Geriatr Psychiatry. 2010 March ; 25(3): 266–274. doi:10.1002/gps.2332.

The Cognitive Change Checklist (3CL): Cross-Validation of a Measure of Change in Everyday Cognition

John A. Schinka^{1,2,3}, Ashok Raj^{2,3}, David A. Loewenstein^{3,4,5}, Brent J. Small^{2,3}, Ranjan Duara^{2,3,4}, and Huntington Potter^{2,3}

¹James A. Haley VA Medical Center, Tampa, FL

²University of South Florida, Tampa, FL

³Florida Alzheimer's Disease Research Center

⁴Wien Center, Miami, FL

⁵University of Miami

Abstract

Objective—The recently developed Cognitive Change Checklist (3CL) is comprised of four scales titled Memory, Executive, Language, and Remote Recall. A previous study demonstrated that the checklist met criteria for reliability and validity in characterizing cases of MCI and early dementia in a memory disorder clinic sample. In this study we examined further the reliability, validity, and efficacy of the 3CL in distinguishing among groups of normal individuals, those with cognitive complaints, amnesic and nonamnesic MCI cases, and early-stage demented individuals.

Design—Scale validation study.

Setting—Alzheimer's Disease Research Center

Participants—One hundred and twenty-five individuals who completed extensive evaluations as part of a longitudinal study of cognitive change.

Results—Scale reliabilities were found to be well within guidelines to support their use in the clinical assessment of change in global and specific cognitive domains. The factor structure of the 3CL was found to be highly similar to that originally reported. Validation support was obtained from correlational analyses that showed significant scale relationships with neurocognitive measures and with MRI ratings of medial temporal atrophy. Informant 3CL scales were found to discriminate individuals with cognitive complaints but without clinical findings from those individuals with amnesic MCI or early dementia.

Conclusion—Results extend the foundation of support for the use of the 3CL as a measure to facilitate identification of cases of MCI and early AD. Further examination of the diagnostic utility of the scale, and of its value in combination with cognitive screening measures, will also be required.

Correspondence: John A. Schinka, Ph.D., James A. Haley VA Medical Center/116B, 13000 Bruce B. Downs Blvd., Tampa, FL 33612, (813) 972-7540, jschinka@health.usf.edu.

No author has a conflict of interest regarding this study.

Keywords

rating scales; everyday function; cognition; aging; cognitive decline; everyday cognition; mild cognitive impairment; dementia Running Header: 3CL Cross-Validation

---

INTRODUCTION

It is now accepted that early diagnosis is a key factor in strategies to ameliorate the economic and social burden of dementia. Unfortunately, while experienced clinicians can reliably diagnose dementias that have progressed to middle stages, the diagnosis of MCI and pre-MCI states continues to be problematic. Longitudinal studies report alarmingly high frequencies of stringently diagnosed research cases of MCI that subsequently revert to normal status. These findings suggest that reliable early diagnosis in clinical and community settings will be difficult to achieve. Even small improvements in the efficacy of any of the components of the diagnostic endeavor (assessment of complaints, reports of IADL/ADL changes, MRI, structured clinical assessment, neuropsychological evaluation) would have a positive effect on reliable identification of early stage cognitive disorders.

The first step in the identification of elderly individuals who are in the early stages of pathological cognitive decline is often a memory concern communicated at a community or clinic-based screening program. While research in aging populations documents declines in cognitive abilities, and an increase in self-reported subjective complaints about loss in cognitive functioning, both cross-sectional and longitudinal studies have failed to show any consistent association between the degree of subjective self-report complaints and objectively measured levels of cognitive performance (Dufouil *et al*, 2005; Frerichs and Tuokko, 2006; Jorm *et al.*, 1994; Jungwirth *et al.*, 2004). In lieu of self-reports, use of informant ratings of cognitive or functional decline has been suggested as an alternative. There is ample evidence to support an informant-based approach in screening and diagnostic efforts (Farias *et al*, 2008; Isella *et al.*, 2006; Jorm, 2004), but existing instruments are not maximally designed to capture recent changes in cognitive status of the types seen in Mild Cognitive Impairment (MCI)(Petersen and Negash, 2008; Petersen *et al.*, 1999).

To facilitate informant evaluations in the earliest stages of cognitive decline, Schinka *et al.* (in press) have developed a brief 28-item measure of cognitive change, the Cognitive Change Checklist (3CL). The 3CL comprises four non-overlapping scales titled Memory, Executive, Language, and Remote Recall. Response options require the patient and/or informant to rate change over the period of the previous two years, using a Likert response scale in which higher scores indicate greater degrees of change. Scale reliabilities were found to be sufficient to support their use in the clinical assessment of change in global cognition (total checklist scale alphas = .96) and specific cognitive domains (individual scale alphas = .84 – .93). Support for the validity of 3CL informant ratings was obtained from correlational analyses that showed significant relationships with neurocognitive measures and from the finding of differences in scale score profiles among diagnostic groups. Most importantly from the standpoint of early diagnosis, significant differences were found between patients with negative evaluations and patients diagnosed as MCI cases for the

informant checklist total, Memory, and Executive scale scores, with effect sizes approaching that found for the MMSE. ROC analyses revealed that the discriminability of the 3CL informant Memory (AUC = .718) and total (AUC = .726) scales was only slightly less than that of the MMSE (AUC = .767).

In this study we describe a validation study designed to examine the psychometric characteristics and validity of the 3CL in a separate, independently diagnosed sample of elderly individuals with heterogeneous cognitive disorders. This sample was well-characterized, both in terms of the comprehensive scope of cognitive performance measures and of MRI ratings of medial temporal atrophy, and allowed us to extend the examination of the efficacy of the 3CL scales by including groups of normals and nonamnesic MCI cases.

METHODS

Study sample

Data for this study were obtained from participants at the Tampa site of a longitudinal study conducted by the Florida Alzheimer's Disease Research Center (FADRC). The focus of the FADRC is the study of cognitive decline in the normal elderly and in individuals with Mild Cognitive Impairment (MCI) or early-stage dementia. FADRC participants are recruited from the community (healthy elderly) and from memory screening programs and memory disorder clinics. Participation is restricted to individuals with an MMSE (Folstein *et al.*, 1976) score greater than 19. The protocol requires that participants complete a baseline evaluation consisting of structured interviews, neurological and mental status exams, an MRI, an extensive battery of cognitive tests, and blood and genetic studies. A subset of 230 individuals were selected to receive a mail packet containing informant and self-report versions of the 3CL together with consent documents and information explaining the study. Instructions directed the participant and informant to complete the 3CL ratings independently, and separate addressed return envelopes with postage prepaid were provided for the participant and informant. This mail procedure had been tested and successfully implemented in a previous study of the 3CL.

Participant and/or informant (93% of whom were spouses) responses were received from 167 (73%) of the individuals receiving mail packets. These ratings were then combined with participant data in the existing ADRC dataset. Of this group, 125 individuals (63 males, 62 females; 90% non-Hispanic Caucasian) had completed the baseline protocol, had no outlying values missing responses on study measures, and received one of the following consensus diagnoses: normal (healthy elderly who respond negatively to a single global item inquiring about memory decline and have an unremarkable diagnostic workup; NORMAL), normal with memory complaint (healthy elderly who respond positively to a single global item inquiring about memory decline and have an unremarkable diagnostic workup; COMPLAINT), Mild Cognitive Impairment (single or multiple domain) with memory being an affected cognitive domain (MCI-A), Mild Cognitive Impairment (single or multiple domain) without memory being affected (MCI-NA), and demented (DEMENTED). The data set for analysis contained self-report ratings for 118 individuals and informant ratings for 85 individuals (7 participants had informant ratings only). This research was completed

with the approval of an Institutional Review Board and all participants were dually enrolled in the FADRC study.

Validity Measures

Measures used to examine the validity of the 3CL scales in correlational analyses included the following cognitive measures: learning (sum of trials 1–3) score and delayed retention percentage (delayed recall score/learning trial 3 score) from the Hopkins Verbal Learning Test-Revised (Benedict *et al.*, 1998), Wechsler Memory Scale-III (Wechsler, 1997a) Visual Reproduction immediate and delayed recall trials; Category Fluency (using the total of animals, vegetables, and fruits), the Mini Mental State Exam (MMSE; Folstein *et al.*, 1975), Boston Naming Test (BNT; Kaplan *et al.*, 1983), Wechsler Adult Intelligence Test-III (Wechsler, 1997b) Digit Span, Digit Symbol, and Block Designs subtests, Controlled Oral Word Association (COWA; Benton and Hamsher, 1978), the Trail-Making Test (Reitan, 1958) Trails A and Trails B times to completion, the Stroop color-word trial (Stroop, 1935), Judgement of Line Orientation (Benton *et al.*, 1983) total correct.

We also examined 3CL correlations with age, total years of education, and MRI ratings of medial temporal atrophy. MRIs were obtained on a 1.5 Tesla MRI machine using proprietary 3-D MP-RAGE (Siemens) or 3-D SPGR (General Electric) sequences. Scans were acquired in the coronal plane and images were reconstructed as 1.5 mm thick contiguous slices. A visual rating system (Urs *et al.*, 2009) was employed by an experienced rater to make bilateral estimates of regional atrophy of brain structures. In previous work, the inter-rater kappa values for the hippocampus (HPC), the entorhinal cortex (ERC), and the perirhinal cortex (PRC) ranged between 0.75 and 0.94, with intra-rater reliabilities ranging from 0.87 to 0.93. The average score for the coronal slices of these three structures at the level of the mammillary body constituted the medial temporal atrophy score for the right (MTA-R) and left (MTA-L) hemispheres, respectively.

Analyses

Descriptive statistics and diagnostic group comparisons were first conducted for demographic variables, MRI atrophy ratings, and neuropsychological measures to insure that the diagnostic groups were defined in a manner consistent with other studies of MCI and dementia.

The limited size of the study sample precluded a confirmatory factor analysis of the four-scale structure of the 28-item 3CL. We instead used an alternative method of examining the factor structure for validation purposes, following a procedure described by Gorsuch (1983). We conducted a principal components analysis that extracted four factors and then calculated the coefficient of congruence (range of possible values is -1.00 to $+1.00$) for each factor with those revealed in the development of the 3CL. To achieve a large coefficient, this procedure thus requires that the full pattern of all 28 item loadings (not just the items with high loadings defining the scale) are similar in both studies. To examine the reliability of the 3CL total scale and individual scales we calculated internal consistency (alpha) reliability coefficients.

The validity of the 3CL was examined by three methods. We first examined the pattern of correlations of 3CL rating with neuropsychological measures, age, years of education, and MRI ratings of medial temporal atrophy. Based on our previous findings, we expected coefficients that showed a more meaningful pattern of validity correlations for informant scales than self-report scales, both in terms of the size of correlations and their pattern with demographic and neuropsychological measures. These analyses excluded the NORMAL group to provide direct comparison with our previous results.

We next examined the 3CL rating profiles of the diagnostic groups, with the expectation that the profile elevations (indicating increasing cognitive problems) would reflect the degree of disease progression in the diagnostic groups. We did not expect to find differences between the NORMAL group and the COMPLAINT group, as cases in the COMPLAINT group were more akin to epidemiologic cases that admit to cognitive changes with age but do not have a concern to the extent that they seek formal screening or treatment. Oneway ANOVAs with diagnostic group as the independent variable were conducted for these analyses, using the F statistic when there were no significant differences in group variances as determined by the Levene test and the Welch (1951) statistic when the Levene test was positive. Tukey HSD tests were used to examine pairwise comparisons between groups following significant F statistics; the Games-Howell (1976) modification of the Tukey HSD test was used following significant Welch statistics.

We conducted ROC analyses of 3CL checklist scale scores to determine the areas under the curve (AUCs) as indicators of the ability of the 3CL to discriminate the MCI and DEMENTED groups from the NORMAL and COMPLAINT groups. Selected neuropsychological measures were included in these analyses to provide comparison data.

RESULTS

Descriptive Statistics

Table 1 presents descriptive data for each of the five diagnostic groups for demographic, MRI, global cognition (MMSE), and neuropsychological test variables. ANOVAs revealed that there were no differences among groups in age ($df = 4, 120, F < 1$) or years of education ($df = 4, 113, F = 1.70, p > .05$). Results for differences in medial temporal atrophy were significant for both right ($df = 4, 46, Welch = 7.56, p < .001$) and left ($df = 4, 46, Welch = 8.02, p < .001$) hemispheres. For the right hemisphere, atrophy ratings were higher for the DEMENTED group than all other groups ($ps < .05$) except for the MCI-A group. The MCI-A group did have higher ratings than the NORMAL and COMPLAINT groups ($ps < .05$), who did not differ from each other. The MCI-NA group only differed from the DEMENT group ($p < .001$). This pattern of results was duplicated for the atrophy ratings of the left hemisphere.

Scores for the MMSE differed significantly among groups ($df = 4, 55, Welch = 16.78, p < .001$). Posthoc comparisons showed that the DEMENTED group scores were lower than the MCI-A ($p < .05$) and NORMAL ($p < .001$) groups. The COMPLAINT group performed lower than the MCI-A ($p < .05$) and DEMENTED group ($p < .001$), but not differently from the NORMAL or MCI-NA groups. MCI-NA group scores differed only from those of the

DEMENT group ($p < .001$). ANOVAs conducted on neuropsychological test scores across groups uniformly revealed significant ($p < .01$) differences. Post-hoc comparisons revealed that for most measures (see Table 1) the NORMAL group performed significantly ($p < .05$) better than the MCI-A and DEMENTED groups, and the MCI-A group performed significantly ($p < .05$) better than the DEMENTED group. The COMPLAINT group generally performed at an equivalent level with the NORMAL group. Findings for the MCI-NA group were more variable.

Factor Analysis and Scale Characteristics

The four factors produced by the principal components analysis with orthogonal (Varimax) rotation were found to have item loadings that were highly correlated with those previously reported. The coefficients of congruence were: Memory, .97; Executive, .94; Language, .95; Remote Recall, .94. Twenty-five of the 28 items had their highest loadings on the expected factor (10/10 for Memory, 5/8 for Executive, 5/5 for Language, 5/5 for Remote Recall). However, the secondary loadings for items not loading most highly on the expected scale exceeded .40 in all cases. Table 2 presents the internal consistency reliability (alpha) coefficients for the 3CL scales for the sample as a whole. The coefficients are slightly higher for the informant version, but all coefficients meet guidelines for reliability in clinical assessment. Table 2 also presents the intercorrelations of the informant and self-report versions. While the within-version correlations generally exceed .70, the same-scale correlations across versions fail to reach a value of .50. All within-version scale intercorrelations exceed the cross-version same-scale correlations.

Correlations with Cognitive Performance Measures

Table 3 presents the correlations of 3CL ratings with medial temporal atrophy ratings and neuropsychological measures. Because of the relatively large sample size, many correlations of small magnitude and little meaningful import were found to be significant. The table therefore presents only correlations equal to or greater than $|\text{.30}|$ (representing 9% shared variance; $p < .01$). No 3CL scales had correlations with age or years of education that were equal to .30. The informant Memory, Executive, and total scale scores were meaningfully correlated with right hemisphere medial temporal atrophy ratings; all informant scales were meaningfully correlated with left hemisphere ratings. No self-report scales were related to atrophy ratings. All 3CL informant scales have moderate correlations with the global measures of cognitive ability, the MMSE. Three of the self-report scales had moderate, but generally lower coefficients.

The majority of 3CL informant score correlations with neuropsychological measures (55 of 60) exceed $|\text{.30}|$ ($p < .01$). In comparison, far fewer of the correlations of self-report 3CL scores with neuropsychological measures exceed $|\text{.30}|$ (21 of 60). In absolute magnitude, the 60 informant coefficients (mean coefficient = .34) are substantively larger than the 60 self-report coefficients (mean coefficient = .19). Within the set of informant intercorrelations, the Memory and Executive scales had strongest associations with learning, memory, and executive cognitive measures, the Language scale had strongest associations with language measures, and the total score had moderate associations across 11 of the 12 cognitive measures.

Comparisons among Diagnostic Groups

Table 4 presents the results of ANOVAs for informant and self-report versions of the 3CL for the five diagnostic groups. ANOVAs for all 3CL scales were significant (see table). The results of posthoc analyses are provided in the table, but the most critical comparisons are reviewed here. None of the self-report scale scores differed significantly between the COMPLAINT and either MCI group. The informant Memory, Executive, and total scale did reveal significantly different scores between the COMPLAINT and MCI-A groups, but no informant scale produced differences between the COMPLAINT and MCI-NA groups. With the exception of the self-report Language scale, all 3CL scales showed significant differences between COMPLAINT and DEMENTED groups.

Paired t-tests were conducted to examine the differences in scores between informant and self-report versions of the 3CL. For the total sample, informant rating scores on the Memory and Executive scales were significantly ($p < .05$) higher than self-report ratings. Differences on the Language and Remote Recall scales were not significant; the difference for total scale scores approached significance ($p = .056$). Sample sizes for individual diagnostic groups were not large enough to conduct sufficiently powered within-group paired t-tests of 3CL scales. For illustrative purposes, however, we present 3CL mean ratings for each group for each scale in Figure 1. The figure contrasts the separation of the MCI-A and DEMENTED groups from the other groups for the informant profiles with the relative lack of profile distinction among the self-report profiles.

Discriminative Ability

ROC curves were generated to measure the effectiveness of selected 3CL scales in classifying COMPLAINT versus MCI and DEMENTED cases. To provide a basis for comparison, ROC curves for the MMSE, the delayed recall trial from Hopkins Verbal Learning Test, and the Digit Symbol test were also produced, as these are commonly used in screening and diagnostic efforts. Results of the ROC analyses are presented in Table 5. For the discrimination of COMPLAINT cases from MCI-A cases, the informant Memory, Executive, and total scales all had AUCs greater than .80 and performed as well or better than the best of the cognitive measures. For the discrimination of the COMPLAINT cases from the MCI-NA cases, only the Digit Symbol test had an acceptable AUC of .72. All measures, including self-report 3CL scales, had at least fair AUCs in discriminating COMPLAINT from DEMENTED cases.

CONCLUSIONS

In this study we attempted to examine further the reliability, validity, and efficacy of the 3CL in a constructive, extended replication of the original research. Alpha coefficients for all scales for both informant and self-report versions of the 3CL were all found to exceed .80, sufficient to support their use in the clinical assessment of global cognition and specific cognitive domains. Examination of the factor structure of the informant version revealed a high level of correspondence with that reported in the checklist's initial development, thus supporting its use across populations with different characteristics.

The results of 3CL correlational analyses with cognitive performance tests were consistent with those previously reported. In comparison with self-report scales, the informant scales had a far more extensive pattern of supportive correlations and a higher mean coefficients with neuropsychological test measures. Notably, we also found a similar pattern for correlations with medial temporal atrophy. To our knowledge, no other informant rating measures have been supported by validation findings with a well-studied biomarker.

Examination of differences for medial temporal atrophy ratings and neuropsychological test scores among the diagnostic groups revealed the expected pattern of results for four of the five groups that composed the overall sample. However, the MCI-NA group only differed significantly from the DEMENTED group in these analyses; no consistent significant differences were found for comparisons with the other groups. These findings were repeated in the analysis of 3CL informant scales. The self-report scales failed to yield differences between complaint and either MCI group. Consistent with previous findings from the initial development study, 3CL informant rating profiles were higher than patient self-report ratings for the sample as a whole; these differences were significant for the Memory and Executive scales. In contrast to previous studies, the self-report rating profile of the DEMENTED patients is higher than those of the other groups. We believe that this reflects that the DEMENTED patients are in the early stage of progression. Inclusion criteria for this ADRC sample required an MMSE score of 20 or greater. Other studies that have found an apparent lack of awareness of cognitive change have typically studied dementia cases in later phases of progression.

We also examined the efficacy of the 3CL by conducting ROC analyses in discriminating COMPLAINT cases from MCI and DEMENTED cases. Our use of COMPLAINT cases rather than NORMALS provided a much more stringent test of 3CL discriminating ability; community-dwelling normals without a cognitive complaint are not appropriate for such comparisons. The difficult clinical task in screening for early-stage cognitive change is to discriminate those with only subjective complaints from true cases of early pathology. Additionally, successful discrimination from MCI cases is most critical; discrimination from progressed cases of dementia is a far less challenging task.

The 3CL Memory, Executive, and total scale scores were found to be as discriminating in distinguishing COMPLAINT from MCI-A cases as the MMSE, Hopkins Verbal Learning Test, or Digit Symbol test, with AUCs that exceeded .80. Similar findings with slightly higher AUCs were obtained for discriminating DEMENTED cases. However, of the 3CL scales and cognitive measures, only the Digit Symbol test marginally separated the COMPLAINT cases from the MCI-NA cases. We are not aware of another study that has examined the efficacy of screening measures for non-amnesic MCI. Our data are somewhat discouraging in that both cognitive measures and 3CL informant measures representing several cognitive domains did not identify the MCI-NA cases. Unfortunately, our sample did not contain enough MCI-NA cases to allow separation into more specific subtypes (e.g., executive) for more detailed analyses.

Despite the overall pattern of positive findings in our analyses, caveats are in order. While our sample did represent heterogeneous diagnoses, it was composed of highly educated

Caucasians. We do not know if our results would generalize to other racial or ethnic groups. Additionally, diagnostic group sample sizes were relatively small, preventing more detailed subgroup analyses and raising the issue of the reliability of estimates of diagnostic efficiency. Future studies should attempt to replicate these findings in Caucasian as well as minority groups, examine normative characteristics of ratings across age and education, and explore the relationship between 3CL ratings and cognitive performance using longitudinal designs. Further examination of the diagnostic utility of the scale, and of its value in combination with cognitive screening measures, will also be required.

ACKNOWLEDGEMENT

This research was supported with the resources and facilities of the James A. Haley Veterans' Hospital and by grant 1P50AG025711-01 from the National Institute on Aging.

REFERENCES

- Benedict RHB, Schretlen D, Goiner L, Brandt J. Hopkins Verbal Learning Test-Revised: Normative data and analysis of inter-form and test-retest reliability. *Clin Neuropsychol.* 1998; 12:43–55.
- Benton, AL.; Hamsher, K. Multilingual Aphasia Examination. Iowa City, IA: University of Iowa; 1978.
- Benton, AL.; Hamsher, K.; Varney, N.; Spreen, O. Judgement of Line Orientation Test. Iowa City, IA: University of Iowa; 1983.
- Dufouil C, Fuhrer R, Alperovitch A. Subjective cognitive complaints and cognitive decline: Consequence or predictor? The Epidemiology of Vascular Aging Study. *J Amer Geriatr Soc.* 2005; 53:616–621. [PubMed: 15817007]
- Farias ST, Mungas D, Reed BR, et al. The measurement of everyday cognition (ECog): Scale development and psychometric properties. *Neuropsychology.* 2008; 22:531–544. [PubMed: 18590364]
- Folstein MF, Folstein SE, McHugh PR, et al. Mini-Mental State. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975; 12:189–198. [PubMed: 1202204]
- Frerichs RJ, Tuokko HA. Reliable change scores and their relation to perceived change in memory: Implications for the diagnosis of mild cognitive impairment. *Arch Clin Neuropsychol.* 2006; 21:109–115. [PubMed: 16216465]
- Games PA, Howell JF. Pairwise multiple comparison procedures with unequal N's and/or variances: A Monte Carlo study. *J Educ Stat.* 1976; 1:113–125.
- Gorsuch, R. Factor Analysis. Mahwah, NJ: Lawrence Erlbaum Associates; 1983.
- Isella V, Villa L, Russo A, et al. Discriminative and predictive power of an informant report in mild cognitive impairment. *J Neurol Neurosurg Psychiatry.* 2006; 77:166–171. [PubMed: 16421116]
- Jorm AF, Christensen H, Henderson AS, et al. Complaints of cognitive decline in the elderly: A comparison of reports by subjects and informants in a community survey. *Psych Med.* 1994; 24:365–374.
- Jorm AF. The Informant Questionnaire on cognitive decline in the elderly (IQCODE): a review. *Int Psychogeriatr.* 2004; 16:275–293. [PubMed: 15559753]
- Jungwirth S, Fischer P, Weissgram S, et al. Subjective memory complaints and objective memory impairment in the Vienna-Transdanube aging community. *Am J Geriatr Soc.* 2004; 52:263–268.
- Kaplan, E.; Goodglass, H.; Weintraub, S. The Boston Naming Test. Philadelphia, PA: Lea & Febiger; 1983.
- Petersen RC, Negash S. Mild cognitive impairment: an overview. *CNS Spectr.* 2008; 13:45–53. [PubMed: 18204414]
- Petersen RC, Smith GE, Waring SC, et al. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol.* 1999; 56:303–308. [PubMed: 10190820]

- Reitan RM. Validity of the Trailmaking Test as an indicator of organic brain damage. *Percept Mot Skills*. 1958; 8:271–276.
- Schinka JA, Brown LM, Proctor-Weber Z. Measuring change in everyday cognition: Development and initial validation of the Cognitive Change Checklist (3CL). *Am J Ger Psychiatry*. (In press).
- Stroop JR. Studies of interference in serial verbal reactions. *J ExpPsychol*. 1935; 18:643–661.
- Urs R, Potter E, Barker W, et al. Visual rating system (VRS) for assessing magnetic resonance images (MRIs): A tool in the diagnosis of MCI and Alzheimer’s Disease. *J Comput Assist Tomogr*. 2009; 33:73–78. [PubMed: 19188789]
- Welch BL. On the comparison of several mean values: An alternative approach. *Biomet*. 1951; 38:330–336.
- Weschler, D. Wechsler Adult Intelligence Scale. 3rd Ed. San Antonio, TX: The Psychological Corporation; 1997.
- Weschler, D. Wechsler Memory Scale. 3rd Ed. San Antonio, TX: The Psychological Corporation; 1997.

Key Points

A recently developed measure of change in everyday cognition was examined in a study designed to investigate its reliability and validity.

The Cognitive Change Checklist (3CL) is a 28-item measure comprised of four scales titled Memory, Executive, Language, and Remote Recall. Self-report and informant versions were administered to individuals who completed extensive evaluations as part of a longitudinal study of cognitive change.

Results extended the foundation of support for the use of the 3CL as a measure to facilitate identification of cases of MCI and early AD.

Figure 1.
Informant and Self-Report 3CL scores for all diagnostic groups.

Table 1

Characteristics of the Diagnostic Groups

Variable	Diagnostic group			
	Normal ^a (n = 27) M ± SD	Complaint ^b (n = 25) M ± SD	MCI-A ^c (n = 25) M ± SD	MCI-NA ^d (n = 27) M ± SD
Age	73.56 ± 6.50	74.04 ± 5.95	75.04 ± 5.71	75.70 ± 4.96
Years of education	15.04 ± 2.29	15.00 ± 3.62	14.00 ± 1.93	13.65 ± 2.06
Atrophy – Right	0.29 ± 0.52 ^{ce}	0.27 ± 0.42 ^{ce}	0.91 ± 0.74 ^{ab}	0.43 ± 0.52 ^e
Atrophy – Left	0.29 ± 0.43 ^{ce}	0.30 ± 0.49 ^e	0.82 ± 0.66 ^a	0.60 ± 0.58 ^e
MMSE	28.96 ± 0.98 ^{ce}	29.20 ± 1.50 ^{ce}	27.54 ± 2.06 ^{abe}	28.04 ± 2.19 ^e
HVLT Trials 1–3	25.15 ± 5.63 ^{ce}	25.64 ± 4.58 ^{cde}	18.29 ± 4.38 ^{abe}	20.63 ± 3.85 ^{abe}
HVLT Retention	0.95 ± 0.18 ^{ce}	0.90 ± 0.26 ^{ce}	0.64 ± 0.32 ^{abe}	0.83 ± 0.20 ^e
Visual Reproduction	31.11 ± 4.36 ^{ce}	30.84 ± 4.73 ^{ce}	22.67 ± 5.48 ^{abde}	28.70 ± 5.93 ^{ce}
Visual Reproduction retention	0.82 ± 0.18 ^{ce}	0.80 ± 0.13 ^{ce}	0.46 ± 0.26 ^{abde}	0.78 ± 0.18 ^{ce}
Trails B time	79.88 ± 25.04 ^{ce}	81.58 ± 29.08 ^e	127.17 ± 57.26 ^{ae}	104.11 ± 48.73 ^e
Digit Symbol	48.63 ± 7.54 ^{cde}	46.52 ± 11.05 ^{de}	38.38 ± 11.08 ^{ae}	38.89 ± 7.47 ^{abe}
Stroop color-word	35.19 ± 8.46 ^{cde}	33.68 ± 7.74 ^{de}	27.63 ± 11.60 ^{ae}	26.37 ± 7.15 ^{abe}
Boston Naming Test	55.00 ± 4.62 ^{cde}	55.08 ± 4.19 ^{de}	48.25 ± 7.63 ^{ae}	51.93 ± 5.38 ^{abe}
Category fluency	47.74 ± 9.09 ^{cde}	46.48 ± 10.67 ^{cde}	34.63 ± 9.18 ^{abe}	38.00 ± 8.47 ^{abe}
Verbal fluency	41.96 ± 9.51 ^{cde}	39.76 ± 7.44 ^{cde}	29.75 ± 14.42 ^{abe}	21.44 ± 5.81 ^{ab}
Judgment of Line Orientation	24.89 ± 4.62	23.68 ± 3.79	23.09 ± 4.50	22.96 ± 4.16
Block Design	33.56 ± 8.77 ^{cde}	34.12 ± 6.90 ^{cde}	26.79 ± 7.63 ^{abe}	26.67 ± 6.80 ^{abe}

Note. Superscripts indicate significant group differences in means ($p < .05$).

Table 2
Scale Alpha Coefficients and Scale Intercorrelations for 3CL Informant and Self-report Versions.

Scale	Number of Items	Alpha	Scale									
			2	3	4	5	6	7	8	9	10	
Informant Version												
1	Memory	28	.97	.89	.73	.77	.96	.34	.41	.32	.27	.37
2	Executive	10	.96		.81	.80	.96	.36	.46	.35	.32	.41
3	Language	8	.94			.69	.86	.37	.46	.49	.36	.45
4	Remote Recall	5	.92				.86	.46	.51	.38	.39	.48
5	Full Scale (Total)	5	.98					.41	.49	.40	.35	.45
Self-Report Version												
6	Memory	28	.93									
7	Executive	10	.88				.82					
8	Language	8	.86					.67	.72			
9	Remote Recall	5	.86					.77	.73	.68		
10	Full Scale (Total)	5	.96					.94	.92	.83	.87	

Note. Coefficients in italics are intraclass correlations; all others are Pearson coefficients. *n* = 85 for Informant Version; *n* = 118 for Self-Report Version.

Table 3

3CL Informant and Self-Report Scale Correlations with Validation Measures.

Measures	3 CL Informant ratings					3 CL Self-report ratings				
	Memory	Executive	Language	Remote	Total	Memory	Executive	Language	Remote	Total
Atrophy – Right	.43	.38			.38					
Atrophy – Left	.49	.52	.37	.42	.51					
MMSE	-.46	-.54	-.31	-.32	-.47	-.31	-.36			-.31
HVLT Trials 1–3	-.51	-.56	-.47	-.41	-.55	-.39	-.41			-.40
Visual Reproduction	-.43	-.44	-.34	-.35	-.44		-.37			
HVLT retention	-.43	-.45	-.35	-.42	-.46	-.40	-.33		-.30	-.36
Visual Reproduction retention	-.47	-.46	-.22	-.33	-.43	-.35	-.34			-.30
Trails B time	.44	.52	.41	.39	.49					
Digit Symbol	-.45	-.52	-.43	-.37	-.49					
Stroop color-word	-.41	-.53	-.48	-.37	-.48					
Boston Naming Test	-.39	-.41	-.51	-.45	-.46	-.31	-.34	-.34		-.35
Category fluency	-.61	-.61	-.58	-.49	-.63	-.34	-.37		-.30	-.37
Verbal fluency	-.31	-.32	-.28		-.31					
Judgment of Line Orientation			-.41							
Block Design	-.40	-.45	-.41	-.38						

Note. Only correlations of .30 and above are presented; all correlations are significant at alpha values less than .05. Participants in the NORMAL group were excluded. n = 66 for Informant Version; n = 94 for Self-Report Version.

Table 4

Descriptive Statistics and Group Comparisons for 3CL Scales

Variable	Normal ^a (n = 27) M ± SD	Complaint ^b (n = 25) M ± SD	MCI-A ^c (n = 25) M ± SD	MCI-NA ^d (n = 27) M ± SD	Demented ^e (n = 21) M ± SD
Informant Version					
Memory	3.84 ± 5.54 ^{bcd}	2.79 ± 3.04 ^{ae}	11.33 ± 7.48 ^a	5.50 ± 5.42 ^a	15.00 ± 10.26 ^{ab}
Executive	1.84 ± 2.73 ^{cde}	1.07 ± 1.64 ^e	5.50 ± 5.60 ^a	2.38 ± 3.72 ^{ae}	10.89 ± 7.17 ^{abd}
Language	1.58 ± 2.24 ^{ce}	1.36 ± 2.62	4.33 ± 4.52 ^a	1.81 ± 2.64	5.28 ± 4.60 ^a
Remote Recall	0.68 ± 1.38 ^e	1.43 ± 2.68 ^e	2.44 ± 3.38	0.94 ± 1.61	5.44 ± 4.71 ^a
Full Scale (Total)	7.95 ± 11.31 ^{bcd}	6.64 ± 9.30 ^{ae}	23.61 ± 18.90 ^a	10.63 ± 11.17 ^{ae}	36.61 ± 24.58 ^{abd}
Self-Report Version					
Memory	1.42 ± 1.69 ^{ee}	3.48 ± 3.03 ^{ee}	5.04 ± 5.60 ^{ab}	4.65 ± 3.93 ^e	10.30 ± 8.01 ^{abd}
Executive	0.75 ± 1.11 ^e	1.76 ± 1.79 ^{ee}	3.04 ± 3.44 ^b	2.12 ± 1.95 ^e	6.30 ± 4.76 ^{abd}
Language	1.04 ± 1.23 ^e	2.52 ± 2.26 ^e	3.48 ± 3.70	2.50 ± 2.42	4.15 ± 2.70 ^{ab}
Remote Recall	0.50 ± 0.78 ^e	0.92 ± 0.95 ^e	2.09 ± 3.45	1.81 ± 2.48 ^e	3.60 ± 3.53 ^{abd}
Full Scale (Total)	3.71 ± 3.67 ^{ee}	8.68 ± 6.95 ^{ee}	13.65 ± 14.72 ^{ab}	11.08 ± 9.02 ^e	24.35 ± 17.30 ^{abd}

Note. Superscripts indicate significant group differences in means ($p < .05$).

Table 5

Results of ROC Analyses.

Variable	Diagnostic Groups									
	COMPLAINT vs MCL-A		COMPLAINT vs MCL-NA		COMPLAINT vs DEMENT					
	AUC	SE	Sig	AUC	SE	Sig	AUC	SE	Sig	
MMSE	0.77	0.09	<.05	0.53	0.11	ns	0.90	0.06	<.01	
HVLT Retention	0.81	0.08	<.01	0.66	0.11	ns	0.87	0.07	<.01	
Digit Symbol	0.66	0.11	ns	0.72	0.10	<.05	0.85	0.07	<.01	
3CL Informant Report										
Memory_P	0.87	0.07	<.01	0.66	0.10	ns	0.83	0.08	<.01	
Exec_P	0.81	0.08	<.01	0.61	0.11	ns	0.88	0.07	<.01	
Lang_P	0.73	0.09	<.05	0.61	0.11	ns	0.80	0.08	<.01	
Remote_P	0.57	0.11	ns	0.49	0.11	ns	0.75	0.09	<.05	
Total_P	0.82	0.08	<.01	0.65	0.10	ns	0.84	0.08	<.01	
3CL Self-Report										
Memory_S	0.62	0.10	ns	0.65	0.10	ns	0.83	0.08	0.00	
Exec_S	0.64	0.10	ns	0.57	0.11	ns	0.82	0.08	0.00	
Lang_S	0.59	0.11	ns	0.50	0.11	ns	0.77	0.08	0.01	
Remote_S	0.57	0.11	ns	0.64	0.11	ns	0.78	0.09	0.01	
Total_S	0.65	0.10	ns	0.60	0.11	ns	0.83	0.08	0.00	

Note. AUC = area under curve. SE = standard error. ns = nonsignificant.