



Published in final edited form as:

Alzheimers Dement. 2014 November ; 10(6): 675–683. doi:10.1016/j.jalz.2013.11.007.

Diagnostic accuracy and practice effects in the National Alzheimer’s Coordinating Center Uniform Data Set neuropsychological battery

Melissa Mathews^{a,d}, Erin Abner^{a,c}, Richard Kryscio^{a,c}, Gregory Jicha^{a,d}, Gregory Cooper^{a,e}, Charles Smith^{a,d}, Allison Caban-Holt^{a,b}, and Frederick A. Schmitt^{a,b,d,*}

^aSanders-Brown Center on Aging, University of Kentucky, Lexington, KY, USA

^bDepartment of Behavioral Science, University of Kentucky, Lexington, KY, USA

^cDepartment of Statistics and Biostatistics, University of Kentucky, Lexington, KY, USA

^dDepartment of Neurology, University of Kentucky, Lexington, KY, USA

^eBaptist Neurology Center, Lexington, KY, USA

Abstract

Introduction—The Uniform Data Set (UDS) neuropsychological battery is frequently used in clinical studies. However, practice effects, effectiveness as a measure of global cognitive functioning, and detection of mild cognitive impairment have not been examined.

Methods—A normative total score for the UDS has been developed. Linear discriminant analysis determined classification accuracy in identifying cognitively normal and impaired groups. Practice effects were examined in cognitively normal and cognitively impaired groups.

Results—The total score differentiates between cognitively normal participants and those with dementia, but does not accurately identify individuals with mild cognitive impairment (MCI). Mean total scores for test-exposed participants were significantly higher than test-naive participants in both the normal and MCI groups and were higher, but not significantly so, in the dementia group.

Conclusion—The total score’s classification accuracy discriminates between cognitively normal versus participants who have dementia. The total score appears subject to practice effects.

Keywords

Practice effects; Longitudinal data; Diagnostic accuracy; Aging; Cognitive testing

1. Introduction

The National Institute on Aging’s (NIA) Alzheimer’s Disease Centers (ADCs) have engaged in comprehensive, multidisciplinary Alzheimer’s research since the 1980s. However, until

2005, individual centers developed their own research protocols, making data sharing somewhat problematic. The Uniform Data Set (UDS) [1] was incorporated into all ADCs in 2005 to standardize data collection across centers and disciplines. This battery was also designed to provide a brief assessment (i.e., 30–45 minutes) of multiple cognitive domains using at least one neuropsychological measure per domain with a target of differentiating between participants with normal cognitive functioning versus Alzheimer's disease (AD) [1]. However, the UDS was not specifically developed to distinguish cases with mild cognitive impairment (MCI) from cognitively normal controls or participants with dementia and may lack the depth and complexity necessary to discern subtle, preclinical cognitive changes.

Normative data for the UDS have been provided by Shirk and colleagues [2] in the form of a web-based calculator that generates z scores for each subtest adjusted for age, gender, and education. Data were provided for individual measures only and issues related to practice effects, global cognitive functioning, longitudinal tracking of cognitive change, and the ability to detect subtle cognitive impairment were not addressed. To optimize the use of cognitive measures in both clinical and research settings, a measure's usefulness in terms of diagnostic discrimination must be evaluated. Although differential diagnosis is routinely and successfully done in traditional neuropsychological clinics with thorough, comprehensive assessment techniques, many researchers seek concise batteries that retain the ability to adequately discriminate between the broad categories of cognitively normal, MCI, dementia, and other neurologic conditions.

Using a single, concise, comprehensive score, as opposed to interpreting performance on individual cognitive tests or cognitive domains, is valued for its simplicity and efficiency. As a result, screening measures, like the Mini-Mental State Exam (MMSE) [3] and the Montreal Cognitive Assessment (MoCA) [4], have become popular methods for screening participants for MCI or dementia. However, such brief screening instruments may not be sufficiently difficult, sensitive, or specific to detect MCI or very mild dementia, especially in the highly educated, high-functioning individuals typically representative of a volunteer research population [5]. In addition, dementia affects most higher order cognitive functions [6,7] to varying degrees, even in the earliest stages. Thus, the development of a composite index of cognition that mitigates ceiling and floor effects typically found with traditional, brief mental status exams may further the purpose of staging and detecting MCI and mild dementia.

There is precedent for combining test scores across multiple procedures to derive a unified total score reflecting global cognitive functioning. Chandler and colleagues [8] developed a total score for the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) battery using a control group of normally aging individuals and a clinical group of participants diagnosed with AD. They further validated the use of the total score for diagnostic purposes in a sample of normal controls and participants with MCI and AD. Chandler and colleagues reported that the total score accurately discriminated between normal cognition and impaired participants (with AD or MCI) and showed high 1-month test-retest reliability and concurrent validity with the MMSE and the Clinical Dementia Rating (CDR) scale [9].

The current study provides a method of determining global cognitive function, discriminating between normal and cognitively impaired groups, and examines the effect of repeated test administrations on longitudinal test data using ADC UDS data from the Sanders-Brown Center on Aging at the University of Kentucky. The total score for the UDS battery was derived from data provided by those participants determined to be cognitively normal at the initial UDS assessment [5,10].

2. Methods

2.1. Study overview

The ADC at Sanders-Brown Center on Aging, University of Kentucky, follows older research volunteers with detailed annual cognitive and clinical assessments, with, in most cases, brain donation at death. Participants may be either cognitively normal or impaired at study entry. Inclusion and exclusion criteria for cognitively normal participants, who enroll in the Biologically Resilient Adults in Neurological Studies (BRAiNS) project, have been described in detail previously [5,10]. Briefly, BRAiNS participants are volunteers ≥ 60 years of age who are free of neurologic disorders, major psychiatric conditions, substance abuse, and significant medical conditions affecting cognition at baseline assessment. All study procedures were approved by the institutional review board of the University of Kentucky, and all participants provided written informed consent. Given that these initially normally aging participants are followed longitudinally until death, cases of MCI and AD naturally developed over time. These participants were followed in a separate cohort until 2005, when the BRAiNS and clinical cohorts were combined under the UDS.

2.2. Participants

Participants in the current analysis included all UK-ADC participants with complete initial UDS assessments ($N = 667$). The UDS total score was developed on a subset of test-naïve participants who were cognitively normal, ≥ 60 years of age, had a CDR Sum of Boxes score ($CDR_{sob} = 0$), MMSE ≥ 25 , and were free from clinically diagnosed cognitive impairment ($n = 250$). The CDR yields 2 scores (i.e., Global Score and Sum of Boxes) and is used to stage dementia severity based on interview responses from patients and informants. The Sum of Boxes score is a total score ranging from 0 to 18 based on the sum of 6 domain scores (i.e., orientation, judgment and problem solving, memory, home and hobbies, personal care, and community affairs) each rated from: normal (0); questionable or very mild dementia (0.5); mild dementia (1); moderate dementia (2); and severe dementia (3). These domains are then combined into a global CDR that ranges from 0 to 3 [11,12].

Because the information from the UDS procedures is used to diagnose participants clinically, for the purposes of group discrimination a coding scheme based on an optimal CDR_{sob} cut score suggested in a recent validation study [13] was used to assign classifications of “normal” ($CDR_{sob} = 0$), “questionable impairment” ($CDR_{sob} = 0.5-2.0$), or “dementia” ($CDR_{sob} > 2$) to the full sample of participants. Questionable impairment is referred to as MCI in what follows. All participants with a $CDR_{sob} = 0$ also received a CDR global score = 0.

2.3. Procedures

All participants completed the UDS neuropsychological measures at baseline. The UDS and its administration have been described in detail by Weintraub and colleagues [14]. Briefly, the currently recommended UDS battery [14] includes the MMSE [3], Wechsler Memory Scale—Revised (WMS-R) Logical Memory IA and IIA [15], WMS-R Digit Span Forward and Backward [15], Category Fluency [7], Boston Naming Test—30 item [16], Wechsler Adult Intelligence Scale—Revised Digit Symbol [17], and Trail Making Test (Trails) Parts A and B [18]. For the current study, 2 additional raw scores were derived: Logical Memory Percent Retention (Logical Memory II \div Logical Memory I) and Trail Making Difference Score (Trails B seconds to complete – Trails A seconds to complete). The percent retention score has been added to later versions of the WMS Logical Memory subtest and serves as an indicator of retention relative to initial encoding. Subtracting the motor speed and visual scanning components of Trails A from Trails B should provide a more accurate assessment of the set-shifting executive component from this task than is typically obtained from the score measuring time to complete Trails B [19,20].

2.4. Score development

Cognitively normal, test-naive participants (n = 250) were used to develop the normative score. Given that each individual test is not scored in the same metric, the first task was to derive a scoring system that would not bias the total score with uneven weighting in favor of those tests yielding a greater number of total points. For example, the Boston Naming Test has a total possible score of 30, whereas Digits Forward has a total possible score of 12, thus differentially weighting the influence of the Boston Naming Test relative to Digits Forward. It was also imperative to retain the clinical meaning and interpretive value of each participant's performance on the individual tests before aggregating them into a total score. Thus, the scoring metric was also designed to capture individual performance on each test relative to the mean of the normative group. Clinically, one's position relative to the mean of a normative group (frequently described in terms of percentile rank) can be diagnostically informative and descriptive labels are often applied relative to one's percentile ranking (Table 1).

Scores ranging from 0 to 6 were assigned to each test based on the participant's percentile ranking relative to the mean of the normal control group (Table 2). Percentile rankings were determined by examining quantile estimates for each subtest generated by the PROC UNIVARIATE command in SAS/STAT (version 9.3). The newly assigned scores reflect the descriptive labeling of cognitive performance typically employed by clinical neuropsychologists. There were 2 instances (i.e., Boston Naming Test and Digits Forward) in which the score ceiling was reached at the 76th percentile (i.e., the high average range of performance); that is, 75% of the scores for both the Boston Naming Test and Digits Forward were below the ceiling score (i.e., 30 and 12, respectively) and ranged from 0 to 29 for the Boston Naming Test and 0 to 11 for Digits Forward. The ceiling scores (i.e., 30 and 12) were obtained by approximately 25% of the participants; thus, a score approximating "superior" performance (92nd percentile) could not be achieved. In these cases, the score was capped at 5 rather than 6. Although the tests are not precisely equal in weight to the other tests in the battery, the difference is 1 point, which leaves little room for undue

influence as compared with untransformed raw scores, which differed greatly from test to test. The UDS total raw (uncorrected) score ranges from 0 to 70. The raw UDS total was then converted to a T score (mean 50, SD 10).

2.5. Statistical analysis

The relationship between the UDS T score and demographic characteristics, age at assessment, years of education, gender, and minority status (white vs. non-white) were investigated using multiple linear regression analysis. A full model containing main effects and 2-way interactions for all variables was initially fit to the data for the test-naive normal group. Nonsignificant independent variables were removed one at a time based on highest P value (starting with interactions) until all remaining variables were significant. Linear discriminant analysis (LDA) was then used to test the UDS T score's ability to correctly classify participants according to the CDR_{sob} score group. The normality of the T -score distribution within score groups was assessed by visual inspection of histograms and with the Kolmogorov-Smirnov test. No obvious violations were detected. Subsequent analyses were carried out to determine whether discriminant ability could be improved with the addition of any combination the demographic variables tested in the regression analysis.

Finally, the effect of familiarity with neuropsychological testing procedures, defined as having any study visits prior to the initial UDS assessment, on mean UDS T scores and classification accuracy is considered. Mean UDS T scores within CDR_{sob} groups were compared across practice groups with t tests. LDA was repeated on the group of participants with previous test exposure to assess the effect of practice on the UDS T score's ability to classify group membership. It is assumed that CDR ratings are independent of history of test exposure. All statistical analyses were performed with SAS/STAT (v9.3) software. Statistical significance was determined at the 0.05 level.

3. Results

3.1. Participants' characteristics

Test-naive participants were comparable to test-exposed participants in education and gender distribution, but were significantly younger in the normal ($P < .0001$) and MCI ($P = .0033$) groups and comprised significantly more minorities (Table 3).

3.2. UDS T-score characteristics

Regression analysis revealed that mean UDS T scores were significantly influenced by all main effects. Age, education, gender, and minority status (all significant at $P < .0001$) affected mean UDS T scores such that participants who were younger, more educated, female, and Caucasian performed better than those who were older, less educated, male, and non-Caucasian, respectively. No interaction terms were significant.

3.3. Classification accuracy

The UDS T score successfully differentiates among the test-naive normal and dementia CDR_{sob} groups, but is markedly less able to identify individuals in the MCI group. LDA with UDS T score as the only explanatory variable incorrectly classified normal group

members as having dementia in 1 of 250 (0.4%) cases and incorrectly classified dementia group members as normal in 2 of 27 (7.4%) cases. By contrast, 55 of 250 (22.0%) normal participants and 5 of 27 (18.5%) dementia cases were classified as MCI. MCI group participants were classified as normal in 19 of 72 (26.4%) cases and as having dementia in 21 of 72 (29.2%) cases. A sensitivity analysis including demographic characteristics as explanatory variables did not improve the classification accuracy. Expanding the questionable impairment group to include CDR_{sob} up to 4.0 did not improve classification accuracy. LDA with UDS *T* score as the only explanatory variable incorrectly classified normal group members as having dementia in 0 of 250 (0.0%) cases and incorrectly classified dementia group members as normal in 1 of 12 (8.3%) cases. By contrast, 51 of 250 (20.4%) normal participants and 1 of 12 (8.3%) dementia cases were classified as MCI. MCI group participants were classified as normal in 21 of 87 (24.1%) cases and as having dementia in 27 of 87 (31.0%) cases.

3.4. Practice effects

Mean UDS *T* scores for test-exposed participants were significantly higher than test-naïve participants in both the normal ($t_{487} = 4.15$, $P < .0001$, Cohen's $d = 0.375$) and MCI ($t_{121} = 2.15$, $P = .033$, Cohen's $d = 0.393$) groups and were higher, but not significantly so, in the dementia group ($t_{51} = 0.78$, $P = .44$, Cohen's $d = 0.214$) (Table 3). These differences were maintained even after mean *T* scores are adjusted for the demographic characteristics of age, education, minority status, and gender (data not shown).

Among participants with previous study visits, LDA with UDS *T* score as the only explanatory variable incorrectly classified normal group members as having dementia in 6 of 239 (2.5%) cases and incorrectly classified dementia group members as normal in 1 of 26 (3.9%) cases. By contrast, 39 of 239 (16.3%) normal participants and 4 of 26 (15.4%) dementia cases were classified as MCI. Compared with the test-naïve group, more MCI group participants were classified as normal (17 of 51 [33.3%]), whereas fewer were classified as having dementia (12 of 51 [23.5%]).

4. Discussion

The UDS has been used at ADCs across the United States since 2005. However, its efficacy as a measure of global cognitive functioning and the effect of repeated administrations on longitudinal test data have remained unexamined. In this study we have described a method of using a total standardized score to distinguish those individuals who are cognitively normal from those who are cognitively impaired. The rationale of such an approach becomes apparent when considering the base rate of individual low test scores in any given population. For example, it has been estimated that anywhere from 22% to 56% of healthy older adults obtain at least one impaired memory test score [21,22], thus reducing the interpretability of single test scores representing cognitive domains in batteries such as the UDS. Given that the UDS uses only 1 or 2 measures for each cognitive domain (e.g., executive function, memory, language, working memory), a global, summary measure of cognition that combines multiple test scores may help to overcome the issues encountered by base-rate low scores.

We found that the UDS *T* score's classification accuracy is adequate for discriminating between test-naive participants who are cognitively normal versus participants who have dementia. Less than 1% of the cognitively normal participants in this sample were misclassified as having dementia, whereas participants with dementia were misclassified as being cognitively normal only 7.4% of the time. However, the UDS *T* score lacks both sensitivity and specificity for detecting MCI. In fact, 55.6% of our MCI sample was misclassified as cognitively normal (26.4%) or demented (29.2%). In a secondary analysis not reported here (available on request), *T* scores for the individual cognitive domains of memory, language, executive function, and working memory were also examined to determine whether a subset of scores was more effective for detecting MCI than the full *T* score. When domain scores were entered in an LDA model, 61.2% of test-naive participants with MCI were misclassified as either cognitively normal (22 of 72, 30.6%) or demented (22 of 72, 30.6%). These findings suggest that the UDS screening battery is not adequate to detect subtle impairments typical of MCI, likely due to the limited assessment of episodic memory. The memory measure used in the UDS battery differs from that used in the CERAD neuropsychological battery, because the UDS uses a paragraph memory measure and CERAD uses a list-learning paradigm. Learning a list of unrelated words may be a more cognitively complex task, and this difference may account for the success of the CERAD battery to adequately detect MCI where the UDS does not. In addition, and perhaps alternatively, MCI manifests in a variety of ways (i.e., amnesic, non-amnesic, single domain, multiple domain) [23], which may lead to high variability in testing performance when considering "MCI" as a homogeneous group. It may be more effective for future studies to consider batteries sufficient to tease out these varied and heterogeneous MCI subtypes.

The importance of detecting early cognitive change is clear based on studies finding that AD-associated neurologic alterations begin many years before an individual becomes fully symptomatic for AD. Although some studies reported that biomarkers such as amyloid- β and tau precede observable clinical changes in cognition, it is important to note that these conclusions are often based on cognitive screening measures similar to the UDS [24–26]. As our results have demonstrated, composite or nonspecific screening measures only detect cognitive impairment after it has substantially progressed. Such brief screening measures often lack sufficient complexity and fail to provide a sustained cognitive challenge to the thinking ability of the participant or patient. This poses a significant challenge for detecting mild dementia in relatively high-functioning individuals who likely have sufficient cognitive reserve to "pass" the screening measures despite experiencing measurable cognitive decline relative to their own baseline abilities [27–29].

It is possible that a more comprehensive, sophisticated battery could unmask early, subtle cognitive changes that occur concurrently with early increases in pathology. In addition, the earliest significant cognitive changes may be observable when one is still performing in the "normal" range, even though they are performing well below their own baseline. Cognitive reserve theory hypothesizes that the brain compensates for damage by using alternative methods of task processing or recruiting undamaged neural networks to perform cognitive tasks. Persons with high premorbid ability (i.e., high cognitive reserve) may be able to

compensate for greater amounts of damage for a longer period of time, especially on simple cognitive screening measures [27,30].

In this study we have also examined practice effects on the UDS *T* score for 3 longitudinally followed participant groups (e.g., cognitively normal, MCI, dementia). To be clear, the study was cross-sectional and addressed differences between test-exposed and test-naive groups. However, the test-exposed group had been followed longitudinally with 1-year repeat assessments with a battery that included many of the same tests included in the UDS (specifically, Logical Memory, Trail Making, Animal Naming, Digit Symbol, and Boston Naming). For interpretations based on longitudinally collected data to be valid, it is important to mitigate systematic, extraneous influences on that data. Unfortunately, the UDS *T* score appears subject to practice effects. Even after adjusting for all demographic variables, scores increased in the test-exposed groups relative to the naive groups, which suggests repeated administrations may lead to increases in the mean scores. This is a striking finding, given that the test-exposed group was significantly older than the test-naive group and that cognitive performance traditionally declines with advancing age. In addition, where other studies have shown a lack of practice effects after short-interval retest periods for persons with MCI [31,32], our findings suggest that UDS *T* scores may significantly increase after 1-year retest intervals, given that the test-exposed group (tested at 1-year intervals) had significantly higher mean scores than the test-naive group. We found that test-exposed participants obtained UDS *T* scores that were, on average, >0.5 standard deviation greater than the test-naive group, approximately 6 *T*-score points. Notably, whereas test-naive MCI-diagnosed participants performed in the impaired range ($T < 35$), the performance of test-exposed participants was low average but still within normal limits. Regarding the effect of practice on classification accuracy, 33.3% of test-exposed participants with MCI obtained a “normal” UDS *T* score compared with 26.4% of the test-naive participants. Not only do participants with MCI show a substantial practice effect on the UDS, but the practice effect likely obscures their overall clinical picture.

The current study possesses several strengths. First, the method of developing the score was designed to minimize differential weighting of individual subtests and preserve the clinical meaning and interpretive value of each subtest in addition to producing a total score that is meaningful in terms of the participant’s performance relative to the sample mean. Second, our findings add to the literature reviewed in the Introduction regarding UDS clinical classification accuracy and UDS test-based practice effects. Mean UDS *T* scores were significantly higher for test-exposed cognitively normal participants and in test-exposed participants with MCI. Scores for dementia participants were also increased, although not significantly. This finding demonstrates a clear need to assess the presence of practice effects in any longitudinal data set using the UDS. In addition, the data provide more evidence that screening measures are insufficient to adequately detect MCI. As the goal of longitudinal dementia research continues to move in the direction of early detection, serious efforts to design a cognitive battery that is adequate for the purposes of early detection will need to be undertaken.

There is an alternative possibility for the finding that the UDS lacks ability to effectively discriminate MCI from normal cognition or dementia. One limitation of the study included

the high variability in the UDS total score observed in the cognitively normal group. Given that diagnostic groups were determined by screening measures that are inherently less sensitive to subtle cognitive decline, such as MMSE total score and CDR Sum of Boxes scores, it is possible that some individuals who were in the beginning stages of cognitive impairment were included in this supposedly “normal” group. In support of this assertion, closer inspection of the UDS data from the individual subtests reveals a number of scores that are substantially below what would be expected from a truly cognitively normal group of individuals. Continued longitudinal follow-up to verify whether these low-scoring participants eventually transition to a diagnosis of MCI or dementia would provide useful information in that regard.

Despite some limitations inherent in all abbreviated batteries, the UDS is used in many sites across the United States; thus, a total score characterizing overall performance may be of use in many ADCs. These tests are also frequently used in private clinics as well; therefore, this score may also be useful in clinical practice. The UDS *T* score shows promise as a tool for discrimination between normal cognitive functioning and dementia. However, future studies would benefit from using more challenging and comprehensive neuropsychological batteries to aid in the detection of subtle cognitive alterations characteristic of MCI. A battery designed to identify and track early cognitive decline will also be necessary for longitudinal studies intending to show response to treatment interventions for MCI. Without more sensitive and reliable methods of detecting early disease states and tracking longitudinal cognitive change, it may be difficult to draw definitive conclusions regarding the diagnosis and treatment of MCI. In addition, conclusions based on results from repeated batteries appear subject to practice effects. As participants appear to gain some benefit from repeated administration, the usefulness of repeated measures for tracking cognitive change is limited unless practice effects can be effectively mitigated.

Acknowledgments

The authors thank the study volunteers, Nancy Stiles, MD, and the clinical core staff at the University of Kentucky Alzheimer’s Disease Center for their invaluable assistance in providing the clinical evaluations. This study was funded by the NIH/NIA (R01AG019241) and the NIA (P30AG028383 and R01AG038651).

References

1. Morris JC, Weintraub S, Chui HC, Decarli C, Ferris S, et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer disease and associated disorders*. 2006; 20:210–216. [PubMed: 17132964]
2. Shirk SD, Mitchell MB, Shaughnessy LW, et al. A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery. *Alzheimers Res Ther*. 2011; 3:32. [PubMed: 22078663]
3. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state.”. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975; 12:189–198. [PubMed: 1202204]
4. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool for Mild Cognitive Impairment. *J Am Geriatr Soc*. 2005; 53:695–699. [PubMed: 15817019]
5. Schmitt FA, Wetherby MMC, Wekstein DR, Dearth CMS, Markesbery WR. Brain donation in normal aging: Procedures, motivations, and donor characteristics from the biologically resilient adults in neurological studies (BRAiNS) project. *Gerontologist*. 2001; 41:716–722. [PubMed: 11723339]

6. Grundman M, Petersen RC, Ferris SH, et al. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch Neurol.* 2004; 61:59–66. [PubMed: 14732621]
7. Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology.* 1989; 39:1159–1165. [PubMed: 2771064]
8. Chandler MJ, Lacritz LH, Hynan LS, et al. A total score for the CERAD neuropsychological battery. *Neurology.* 2005; 65:102–106. [PubMed: 16009893]
9. Berg L. Clinical Dementia Rating (CDR). *Psychopharmacol Bull.* 1988; 24:637–639. [PubMed: 3249765]
10. Schmitt FA, Nelson PT, Abner E, et al. University of Kentucky Sanders-Brown Healthy Brain Aging Volunteers: donor characteristics, procedures, and neuropathology. *Curr Alzheimer Res.* 2012; 9:724–733. [PubMed: 22471862]
11. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *Int Psychogeriatr.* 1997; 9(Suppl. 1):173–176. [PubMed: 9447441]
12. Williams MM, Storandt M, Roe CM, Morris JC. Progression of Alzheimer's disease as measured by Clinical Dementia Rating Sum of Boxes scores. *Alzheimers Dement.* 2013; 9(Suppl):S39–S44. [PubMed: 22858530]
13. O'Bryant SE, Lacritz LH, Hall J, et al. Validation of the new interpretive guidelines for the Clinical Dementia Rating Scale Sum of Boxes Score in the National Alzheimer's Coordinating Center Database. *Arch Neurol.* 2010; 67:746–749. [PubMed: 20558394]
14. Weintraub S, Salmon D, Mercaldo N, et al. The Alzheimer's Disease Centers' Uniform Data Set (UDS): the neuropsychologic test battery. *Alzheimer Dis Assoc Disorders.* 2009; 23:91–101.
15. Wechsler, D.; Stone, CP. Manual: Wechsler Memory Scale. New York: Psychological Corporation; 1973.
16. Saxton J, Ratcliff G, Munro CA, et al. Normative data on the Boston Naming Test and two equivalent 30-item short forms. *Clin Neuropsychologist.* 2000; 14:526–534.
17. Wechsler, D. Manual: Wechsler Adult Intelligence Scale. New York: Psychological Corporation; 1955.
18. Armitage SG. An analysis of certain psychological tests used for the evaluation of brain injury. *Psychol Monogr.* 1946; 60:1–48.
19. Lamberty GJ, Putnam SH, Chatel DM, Bieliauskas LA, Adams KM. Derived Trail Making Test Indexes—a preliminary report. *Cogn Behav Neurol.* 1994; 7:230–234.
20. Strauss, E.; Sherman, E.; Spreen, O. A compendium of neuropsychological tests: Administration, norms, and commentary, third edition. Oxford: Oxford University Press; 2006.
21. Brooks BL, Iverson GL, Holdnack JA, Feldman HH. Potential for misclassification of mild cognitive impairment: a study of memory scores on the Wechsler Memory Scale-III in healthy older adults. *J Int Neuropsychol Soc.* 2008; 14:463–478. [PubMed: 18419845]
22. Brooks BL, Iverson GL, White T. Substantial risk of "accidental MCI" in healthy older adults: base rates of low memory scores in neuropsychological assessment. *J Int Neuropsychol Soc.* 2007; 13:490–500. [PubMed: 17445298]
23. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med.* 2004; 256:183–194. [PubMed: 15324362]
24. Landau SM, Mintun MA, Joshi AD, et al. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Ann Neurol.* 2012; 72:578–586. [PubMed: 23109153]
25. Buchhave P, Minthon L, Zetterberg H, Wallin AK, Blennow K, Hansson O. Cerebrospinal fluid levels of beta-amyloid 1–42, but not of tau, are fully changed already 5 to 10 years before the onset of Alzheimer dementia. *Arch Gen Psychiatry.* 2012; 69:98–106. [PubMed: 22213792]
26. Terry RD, Masliah E, Salmon DP, et al. Physical basis of cognitive alterations in Alzheimer's disease—synapse loss is the major correlate of cognitive impairment. *Ann Neurol.* 1991; 30:572–580. [PubMed: 1789684]
27. Stern Y, Albert S, Tang MX, Tsai WY. Rate of memory decline in AD is related to education and occupation—cognitive reserve? *Neurology.* 1999; 53:1942–1947. [PubMed: 10599762]

28. Stern Y, Alexander GE, Prohovnik I, Mayeux R. Inverse relationship between education and parietotemporal perfusion deficit in Alzheimer's disease. *Ann Neurol.* 1992; 32:371–375. [PubMed: 1416806]
29. Stern Y. Cognitive reserve. *Neuropsychologia.* 2009; 47:2015–2028. [PubMed: 19467352]
30. Welsh-Bohmer KA, Ostbye T, Sanders L, et al. Neuropsychological performance in advanced age: influences of demographic factors and apolipoprotein E: findings from the Cache County Memory Study. *Clin Neuropsychologist.* 2009; 23:77–99.
31. Duff K, Beglinger L, van der Heiden S, et al. Short-term practice effects in amnesic mild cognitive impairment: implications for diagnosis and treatment. *Int Psychogeriatr.* 2008; 20:986–999. [PubMed: 18405398]
32. Cooper DB, Lacritz LH, Weiner MF, Rosenberg RN, Cullum CM. Category fluency in mild cognitive impairment—reduced effect of practice in test-retest conditions. *Alzheimers Dis Assoc Dis.* 2004; 18:120–122.

RESEARCH CONTEXT

1. **Systematic Review:** We reviewed literature pertaining to practice effects, diagnostic discrimination, and constructing summary scores. Specifically, we reviewed literature pertaining to the CERAD neuropsychological battery as the total score constructed from this battery was influential in constructing the UDS total score. Findings related to practice effects in older adults and in impaired populations were also reviewed.
2. **Interpretation:** The information in this study will benefit clinicians and researchers conducting longitudinal, clinical trials. Our findings are significant because they demonstrate practice effects in a widely used neuropsychological battery. We also demonstrate that the UDS neuropsychological battery effectively discriminates between cognitively normal and participants with dementia, although it is not sensitive to the subtle changes seen in mild cognitive impairment.
3. **Future Directions:** We hope to make investigators aware of this phenomenon and encourage the discovery of new ways to address practice effects and case ascertainment issues in clinical trials.

Table 1

Percentile rankings and clinical descriptors

Percentiles	Clinical descriptors
0–2	Impaired
3–8	Borderline
9–24	Low average
25–50	Average (bottom half of average range)
51–75	Average (top half of average range)
76–91	High average
92+	Superior

Table 2

Conversion of subtest raw scores to new scores based on percentile rank

Test	Subtest raw score	Percentiles	New score	Descriptor
Logical Memory I	0-3	0-2	0	Impaired
	4-6	3-8	1	Borderline
	7-9	9-24	2	Low average
	10-12	25-50	3	Average -
	13-15	51-75	4	Average +
	16-17	76-91	5	High average
	18+	92+	6	Superior
Logical Memory II	0-2	0-2	0	Impaired
	3-5	3-8	1	Borderline
	6-8	9-24	2	Low average
	9-11	25-50	3	Average -
	12-13	51-75	4	Average +
	14-16	76-91	5	High average
	17+	92+	6	Superior
Logical Memory Retention	0.0-0.39	0-2	0	Impaired
	0.4-0.57	3-8	1	Borderline
	0.58-0.76	9-24	2	Low average
	0.77-0.90	25-50	3	Average -
	0.91-1.00	51-75	4	Average +
	1.10-1.12	76-91	5	High average
	1.13+	92+	6	Superior
Animals	0-8	0-2	0	Impaired
	9-13	3-8	1	Borderline
	14-15	9-24	2	Low average
	16-19	25-50	3	Average -
	20-23	51-75	4	Average +
	24-27	76-91	5	High average
	28+	92+	6	Superior
Vegetables	0-7	0-2	0	Impaired
	8-9	3-8	1	Borderline
	10-12	9-24	2	Low average
	13-14	25-50	3	Average -
	15-18	51-75	4	Average +
	19-20	76-91	5	High average
	21+	92+	6	Superior
Boston	0-20	0-2	0	Impaired
	21-23	3-8	1	Borderline
	24-25	9-24	2	Low average
	26-28	25-50	3	Average -

Test	Subtest raw score	Percentiles	New score	Descriptor
Digit Span Backward	29	51–75	4	Average +
	30	76–91	5	High average
	0–3	0–2	0	Impaired
	4	3–8	1	Borderline
	5	9–24	2	Low average
	6–7	25–50	3	Average –
	8	51–75	4	Average +
Digit Span Forward	9–10	76–91	5	High average
	11+	92+	6	Superior
	0–5	0–2	0	Impaired
	6	3–8	1	Borderline
	7	9–24	2	Low average
	8–10	25–50	3	Average –
	11	51–75	4	Average +
Digit Symbol	12	76–91	5	High average
	0–22	0–2	0	Impaired
	23–28	3–8	1	Borderline
	29–35	9–24	2	Low average
	36–45	25–50	3	Average –
	46–52	51–75	4	Average +
	53–61	76–91	5	High average
Trails A	62+	92+	6	Superior
	102+	0–2	0	Impaired
	101–69	3–8	1	Borderline
	68–50	9–24	2	Low average
	49–39	25–50	3	Average –
	38–31	51–75	4	Average +
	30–25	76–91	5	High average
Trails B	>25	92+	6	Superior
	274+	0–2	0	Impaired
	273–171	3–8	1	Borderline
	170–112	9–24	2	Low average
	111–88	25–50	3	Average –
	87–66	51–75	4	Average +
	65–53	76–91	5	High average
Trails Difference	>53	92+	6	Superior
	171+	0–2	0	Impaired
	170–126	3–8	1	Borderline
	125–69	9–24	2	Low average
	68–47	25–50	3	Average –
	46–32	51–75	4	Average +
	31–18	76–91	5	High average

Test	Subtest raw score	Percentiles	New score	Descriptor
	<18	92+	6	Superior

Table 3

UK-ADC Participants' Characteristics at Initial UDS Assessment (N = 667)

	Normal (CDR _{sob} = 0)		MCI (0.5 CDR _{sob} 2.0)		Dementia (CDR _{sob} > 2.0)	
	Test naive (n = 250)	Test exposed (n = 239)	Test naive (n = 72)	Test exposed (n = 51)	Test naive (n = 29)	Test exposed (n = 26)
Age, years (mean ± SD)	72.7 ± 6.0	75.4 ± 7.2*	76.0 ± 6.5	80.0 ± 7.9*	77.6 ± 6.2	79.6 ± 8.3
Education, years (mean ± SD)	16.1 ± 2.9	16.1 ± 2.4	16.0 ± 3.5	15.9 ± 3.0	14.5 ± 4.2	14.8 ± 3.9
Minority (%)	21.2	1.7 [†]	31.9	5.9 [†]	31.0	19.2
Female (%)	66.4	65.7	44.4	51.0	62.1	50.0
UDS raw total (mean ± SD)	41.7 ± 8.9	45.0 ± 8.8*	26.7 ± 11.7	31.3 ± 11.7*	13.0 ± 10.1	15.3 ± 10.4
UDS T score (mean ± SD)	50.0 ± 10.0	53.7 ± 9.9*	33.1 ± 13.1	39.3 ± 13.3*	17.8 ± 11.4	20.3 ± 11.7

* Significantly greater than test naive ($P < .05$).

[†] Significantly less than test naive ($P < .05$).