Published in final edited form as:

J Appl Gerontol. 2014 December ; 33(8): 942–962. doi:10.1177/0733464812455099.

Validation of a Self-Administered Computerized System to Detect Cognitive Impairment in Older Adults

Samuel D. Brinkman¹, Robert J. Reese¹, Larry A. Norsworthy³, Donna K. Dellaria¹, Jacob W. Kinkade¹, Jared Benge⁴, Kimberly Brown⁵, Anna Ratka⁶, and James W. Simpkins⁵

¹Neuropsychology Clinic, Abilene, TX

²University of Kentucky, KY, USA

³Abilene Christian University

⁴Jack C. Montgomery VA Medical Center, Muskogee, OK

⁵University of North Texas Health Science Center

⁶Texas A&M Health Science Center

Abstract

There is increasing interest in the development of economical and accurate approaches to identifying persons in the community who have mild, undetected cognitive impairments. Computerized assessment systems have been suggested as a viable approach to identifying these persons. The validity of a computerized assessment system for identification of memory and executive deficits in older individuals was evaluated in the current study. Volunteers ($N = 235$) completed a 3-hr battery of neuropsychological tests and a computerized cognitive assessment system. Participants were classified as impaired ($n = 78$) or unimpaired ($n = 157$) on the basis of the Mini Mental State Exam, Wechsler Memory Scale-III and the Trail Making Test (TMT), Part B. All six variables (three memory variables and three executive variables) derived from the computerized assessment differed significantly between groups in the expected direction. There was also evidence of temporal stability and concurrent validity. Application of computerized assessment systems for clinical practice and for identification of research participants is discussed in this article.

Keywords

Dementia; Alzheimer's disease; screening; memory; executive functions; delayed alternation task; diagnosis; mild cognitive impairment

© The Author(s) 2012

Reprints and permissions: sagepub.com/journalsPermissions.nav

Corresponding Author: Robert J. Reese, University of Kentucky, 235 Dickey Hall, Lexington, KY 40514, USA. jeff.reese@uky.edu.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Coauthor, Samuel D. Brinkman, PhD, is President of Dementia Screening, Inc., developer of the GrayMatters® system.

There is increasing interest in developing cost-effective approaches to early identification of older persons with cognitive decline. However, there is a lack of consensus at this time regarding whether, when, and how screening for cognitive deficits should take place (Ashford et al., 2006). There are a number of reasons for this, including the lack of more effective treatments which can be used early in the diagnosis, the unavailability of screening procedures that provide acceptable accuracy at reasonable costs, and lack of empirical evidence that screening improves patient management or clinical outcome (Boustani, Peterson, Hanson, Harris, & Lohr, 2003). Nevertheless, accurate, low-cost screening continues to be a very attractive clinical goal (Ganguli, 1997; Snitz, Morrow, Rodriguez, Huber, & Saxton, 2008; Wild, Howieson, Webbe, Seelye, & Kaye, 2008).

It is estimated that 5.3 million Americans have Alzheimer's disease (AD), and this number is projected to increase markedly over the next few decades. By 2029, there will be an estimated 70 million people in the United States who are at least 65 years old and consequently at increased risk for developing AD or other Dementia (Alzheimer Association, 2009). A number of drugs for AD are currently in Phase II and Phase III clinical trials. These drugs are thought to modify both the pathological and the temporal course of the disease. The availability of such drugs will place increasing emphasis on effective screening so that the benefits of the drugs (both to the individual and to the health care system) can be realized. Unfortunately, a patient's cognitive deficits have often progressed for several years at the time of diagnostic work-up and initiation of treatment. As a result, the patient may be too impaired to respond optimally to medications (Boustani et al., 2005; Guilford & Cummings, 1999).

These clinical considerations and uncertainties make the economic affect of screening and medications difficult to estimate (Jonsson, 2003). The annual cost of AD to the U.S. economy (including government-funded programs and private industry), however, is thought to be at least in the tens of billions of dollars (Bloom, de Pouvoirville, & Straus, 2003). Weimer and Sager (2009) have estimated the degree to which screening and early diagnosis may result in net social benefits as well as financial savings for state and federal governments. These authors cited evidence that medication treatments (central acetyl cholinesterase [Ach] inhibitors) and nonmedication treatments (caregiver support) each independently reduce the risk of nursing home placement and consequently may result in financial savings for state and federal governments. Applying a model in which the rate of progression of Dementia is decreased through appropriate medication use and caregiver support, the authors assessed the relative savings when treatment onset occurred early in the clinical progression as opposed to later in the progression of the disease. Their findings demonstrated a very significant financial benefit that may be achieved as a result of early diagnosis and consequent early intervention. In order to maximize clinical, financial, and personal outcomes that may result from improvements in Dementia treatments, effective and reliable methods of screening for early symptoms of Dementia must be available.

Unfortunately, commonly used instruments for Dementia screening are personnel intensive and consequently expensive to apply. The cost of such approaches necessarily limits their application to screening efforts when large groups of individuals are targeted. The Mini-Mental State Examination (MMSE) has been the most widely used screening instrument for

cognitive impairments in older individuals, but a number of other approaches are available as well (Folstein, Folstein, & Fanjiang, 2001; Lorentz, Scanlan, & Borson, 2002). These approaches require trained professionals and involve evaluation of patient performance in various tasks. Tornatore, Hill, Laboff, and McGann (2005) have suggested that screening might be accomplished more economically using self-administered computerized systems which employ a touch-screen monitor to reduce the complexity of patient interactions with the computer. These authors evaluated 310 participants in Washington State and classified participants as Normal or Mild Cognitive Impairment (MCI) on the basis of the Wechsler Memory Scale-Revised Logical Memory Subtest delayed recall score (cutoff was the 10th percentile). They demonstrated that meaningful clinical data can be derived for older participants when a self-administered computerized program with touch-screen interface is used. Their data support the use of this approach when screening for cognitive impairment in older adults.

The application of computer technologies offers several key advantages over conventional assessment approaches (Wild et al., 2008). First, the patient may be able to interact with the computer without the continuous attendance of highly trained staff and the consequent personnel expenses (*self-administered* screening), thus conserving financial resources. Second, well-designed computer programs can administer even complex or timed tasks in an accurate, consistent, and standardized manner. Third, the computerized approach can make scoring, data handling, and reporting automated and therefore more economical. Computerized applications also have the advantage of an easy interface with paperless record-keeping systems.

The use of touch-screen technology for screening requires the development of tasks amenable to this evaluation approach and sensitive to the domains that demonstrate decline in early stages of AD. It has been well-established that early deficits in AD involve memory abilities and executive functions (Albert, Moss, Tanzi, & Jones, 2001; Baddeley, Baddeley, Bucks, & Wilcock, 2001; Chenet et al., 2000; Cahn et al., 1995; Duke & Kaszniak, 2000; Linn et al., 1995; Perry & Hodges, 1999; Sgaramella et al., 2001). Although traditional tests of these functions, such as verbal memory tests with free recall, are the standard against which newer approaches should be measured, they have not been adapted to self-administered computerized administration. Thus, there is a need for new paradigms and tasks which use the strengths of the computerized administration approach while providing information in key domains of interest.

The current study addresses the issue of whether the Visual Delayed Recognition (VDR) task and Delayed Alternation Task (DAT), as parts of a proprietary computerized system called GrayMatters®, can be useful tools for monitoring key cognitive domains in older individuals. The first step in this process is to determine whether older individuals who do not have evidence of cognitive impairment perform reliably better on these tasks than older individuals who demonstrate cognitive impairments on more established psychometric instruments but who do not have a diagnosis of Dementia.

There are other computerized assessments that already exist and offer promise, and may lead one to ask if another is needed. We believe so for three reasons. First, most computerized

assessments lack the empirical foundation (e.g., normative data, adequate psychometric studies) or increased feasibility over traditional assessment methods (e.g., can be self-administered, administration time is less than 30 min). Tierney and Lermer (2010) identified 11 computer-based assessments to identify cognitive impairment but surmised only three had the requisite research and efficiency to merit being recommended. Second, more research is needed to identify which assessment methods and tasks are most effective in an electronic format. For example, Scharre et al. (2010) indicated that computerized assessments emphasize executive functioning or memory tasks to differing degrees. Third, the assessment evaluated in the current study offers advantages that many other computerized assessments do not, including the presentation of directions both visually and orally and only taking 20 min to complete.

For the current study, we had three hypotheses. First, we hypothesized that participants identified as cognitively impaired would receive significantly lower scores on the VDR task and DAT than nonimpaired participants. Second, we hypothesized that the VDR and DAT scores would generate temporally stable scores (one month). Third, we hypothesized that the VDR and DAT scores would evidence concurrent validity with established measures used to identify cognitive impairment.

The VDR task was designed to measure visual memory. There is substantial evidence that indicates visual-based memory tasks are effective for identifying cognitive decline in older adults and subsequent AD (Alladi, Arnold, Mitchell, Nestor, & Hodges, 2006; Blackwell et al. 2004; Haught, Weber, Demarest, Keefover, & Rankin, 1996; Lee et al., 2010; Saunders & Summers, 2011). Moreover, there is some evidence that suggests the use of visual memory tasks may even be more sensitive for identifying MCI than verbal memory tasks (e.g., Alladi et al., 2006).

The DAT paradigm, originally described by Hunter (1913), may be useful as a measure of executive functioning in computerized applications. Several studies have suggested that the DAT paradigm is a useful measure of executive functioning (Archibald & Kerns, 1999; Oscar-Berman & Zola-Morgan, 1980; Oscar-Berman, Zola-Morgan, Oberg, & Bonnder, 1982; Partiot et al., 1996) and that the paradigm has applicability to early recognition of AD (Collins, 2000). Bengé (2003) demonstrated that the computerized version used in this study correlated well with established measures of executive functions in a sample of cognitively normal older adults.

Method

Measures

GrayMatters® visual delayed recognition task (VDR)—Visual memory was measured with the VDR task, a computerized forced-choice format for measuring the ability to acquire and retain new visual information. Pictures of objects are presented to the participant while the participant is cued verbally to study the pictures. The pictures are removed, and a picture is presented to the participant while he/she is asked (with verbal and written cues) whether the picture (*challenge* picture) was one of those just presented. The participant is instructed to touch either the *Yes* button or the *No* button on the touch-screen

monitor. There are 12 trials. Trial one was designed to have a very high probability of correct responding (based on pilot testing with young adult normal participants), with only two pictures presented simultaneously for 5 sec and challenge items presented after a 5-sec delay. Subsequently 11 additional sets of pictures are presented, and each of these picture sets consists of four pictures. After a 5-sec delay, two challenge pictures are presented (sequentially) after each set of pictures and the participant responds as above. There is a 50/50 ratio of correct to incorrect challenge items. Trials 5 to 8 and 9 to 12 are equated in terms of difficulty level of the pictures (based on evaluation of a much larger pool of potential items, using 100 young adult participants and determining probability of pass or failure with the items). On trials 9 to 12, however, the participant is prompted to complete several relatively simple distracter tasks (such as identifying shapes) during the 10-sec delay interval to decrease the opportunity to rehearse. The distracter tasks are also completed on the computer screen using touch-screen responses. Scores from the VDR are the total number of correct responses on the challenge trials, the number of false positive errors, and the number of correct responses on distracter tasks.

GrayMatters® delayed alternation task (DAT)—The DAT consists of a problem-solving task with paradigm shifts. Four tasks, each using a different rule, are presented sequentially to the participant. First, pictures of two hands are presented on the touch-screen monitor. The pictures are actually mirror images of the same hand to ensure that the hands are visually identical. The participant is asked to touch the hand under which would be found a coin. When the participant touches the screen, the hand image is turned over to reveal either the coin or an empty hand. After a 10-sec delay (in keeping with the classic DAT paradigm), the hands are again presented and the participant is again prompted visually and verbally to select the hand which has the coin. The visual prompt is text presented on the screen and the same information is given via a verbal prompt provided by the computer-based program. Under this delayed alternation (DA) rule, the coin is *placed* in the opposite hand after a correct response and remains in the same hand after an incorrect response. Response on the first trial of the DA rule is correct regardless which hand was selected. Twenty-five individual trials are completed unless the participant selects the correct hand 5 trials in a row (success criterion reached).

After 25 trials (failure criterion) or attainment of the success criterion, the participant begins the second task that uses the nonalternating rule (NA). Under this rule, the coin remains in the same hand over all the trials. Again the first trial is successful regardless which hand was selected. The same success and failure criteria are used. The 10-sec delay period between trials is discontinued under the NA and subsequent rules to decrease the overall assessment time, as the primary interest in the NA and subsequent rules is to assess ability to shift cognitive set following the DA task.

For the third task, the hands are replaced by shapes: a blue circle and a red square. The third rule is termed the Shape Alternation (ShA) rule. The participant is asked to select which shape is covering the coin. Response on the first trial is always correct, and the coin is subsequently changed to the opposite shape after a correct guess, whereas it remains under the same shape after incorrect guesses. The locations (right or left side of screen) of the two

shapes vary according to a predetermined sequence during each of the 25 trials. The same success and failure criteria are used.

The fourth task uses the same shapes from the previous task for the Side Alternating (SA) rule. Under the SA rule, the coin alternates to the opposite side after each correct response and remains on the same side after incorrect responses. Again the locations of the two shapes vary in a predetermined sequence for each of the 25 trials. The same success and failure criteria are applied.

Scores from the DAT are the number of rules for which the success criterion is met, total number of correct responses on the four rules, and the number of perseverative errors. A perseverative error is defined as errors on consecutive responses.

Cognitive measures—For the purpose of initially identifying participants who were cognitively impaired, a battery of established cognitive measures was used. All participants completed the Mini Mental State Exam (MMSE; Folstein, Folstein, & Fanjiang, 2001), the Trail Making Test, Part B (TMT; Reitan & Wolfson, 1985), and the Wechsler Memory Scale-Third edition (WMS-III; Wechsler, 1997). The MMSE was selected because it is a widely used screen for cognitive deficits, and the TMT and WMS-III were selected because they are well-established measures for evaluating cognitive issues specific to executing functioning and memory, respectively.

Procedure

Participants ($N = 235$) were recruited through public media (newspaper ads, public service announcements, pamphlets and postings at relevant centers, and letters to local physicians offering cognitive screening services) and from persons who came to the Neuropsychology Clinic in Abilene, TX (either as clinical referrals or as family members or caretakers of patients). The recruitment procedures offered free evaluations of memory for individuals

50 years of age. Participants were recruited in 11 cities in Texas. Participants recruited at one site, University of North Texas—Health Sciences Center (UNTHSC), were given a small amount of money to offset travel expenses (US\$10.00 each). Exclusion criteria included a diagnosis of MCI or Dementia prior to the study, history of a relevant and confounding neurological disorder (such as stroke, traumatic brain injury with residual known deficits, seizure disorder, etc.), sensory or motor loss sufficient to impair performance on the assessment procedures, current potentially confounding systemic illness, or severe psychiatric disorder. For all participants this was their first exposure to the self-administered GrayMatters® system.

Procedures performed in this clinical study were approved by the Institutional Review Board (IRB) at the UNTHSC in Fort Worth. The following data collection procedures were performed sequentially after informed consent was obtained: A standardized clinical interview by a member of the research team that obtained each participant's medical, neurological, and psychiatric history and any medications being taken (30–60 min); administration of standardized assessment procedures (MMSE, TMT, and WMS-III); and completion of the computerized GrayMatters® assessment (VDR and DAT). The VDR and DAT assessments were completed on a desktop computer using 15.1" touch-screen monitors

and digitized voice instructions presented over the built-in speakers. The entire battery of tests took approximately 3 hr to administer. Participants were also asked at initial screening whether they would be willing to repeat the computerized assessments again in 1 month. Ninety participants agreed to the follow-up and completed the computerized assessment again (no statistically significant differences on any of the measures were found between those who followed up and those who did not). If desired and authorized by the participant, results of the MMSE, WMS-III, and TMT were sent to the participant's primary care physician (PCP).

Data collection took place at long-term care (LTC) facilities and at senior citizen centers in the various cities. Acceptable arrangements were made to secure participants' privacy in a relatively nondistracting environment for data collection at each site. An assistant was present to help if participants incurred problems, but assistance was rarely requested (<5). Responses for the VDR and DAT were collected on the GrayMatters® computer and later transmitted automatically to a central server (using proprietary communications protocols) for storage and later statistical analysis. Patient data were collected and stored in compliance with HIPAA procedures at the Neuropsychology Clinic.

The clinical data (including data taken from history, MMSE, TMT, and WMS-III) were evaluated without knowledge of the results of the GrayMatters® evaluation and participants was classified as unimpaired or impaired based on the neuropsychological test performance and MMSE as evaluated by the first author, a licensed neuropsychologist. Participants were considered impaired if they had an MMSE score ≤ 25 or if there was evidence based on the neuropsychological battery of tests that memory or executive functions were in the impaired range. Age-adjusted norms were used for the neuropsychological tests. As the MMSE has a very high specificity, a score ≤ 25 was accepted as evidence of impairment. If the MMSE score was higher than 25, however, the participant was considered impaired if the neuropsychological test data were consistent with impairment. In other words, if the MMSE score was in the normal range, a participant might still be found to be impaired based on the more detailed and sensitive neuropsychological battery. Impairment on the neuropsychological tests was defined as:

- a. Trail Making Test B performance (time and error scores) below 10th percentile according to norms of Ashendorf et al. (2008) or,
- b. At least two of five WMS-III Index scores (at least one of which must be a delayed recall or recognition recall score) ≥ 1.5 standard deviations below mean.

Data Analysis Plan

Descriptive statistics for demographic variables and group comparisons for the unimpaired and impaired groups were completed using parametric and nonparametric procedures where appropriate. The GrayMatters® variables were evaluated in terms of distributional characteristics, and group comparisons were completed for the various VDR and DAT scores using one-way multivariate analysis of covariance (MANCOVA), controlling for age and education, independent *t*-tests, while change in scores at retesting were determined using within-groups *t*-tests. Because the distributions for DAT data were skewed,

nonparametric statistics were used with these data (Spearman's Rho for correlations, Mann-Whitney *U* test and Wilcoxon Signed Ranks Test for independent and dependent group comparisons, respectively). We used the Bonferroni method to guard against inflated Type I error rates because of the six univariate follow-up comparisons conducted; therefore, our adjusted *a* for these comparisons is .008. Bivariate correlations were computed to evaluate the test-retest for the VDR and DAT as well as concurrent validity with the established neuropsychological tests.

Results

Sample Characteristics

All 235 participants recruited into the study were able to complete the testing. Most of the participants were Caucasian ($n = 222$) with five participants being African American (*4 unimpaired, 1 impaired*) and eight Hispanic (*4 unimpaired and 4 impaired*). There were no statistically significant differences between racial/ethnic groups on any of the five measures (described in Measures section). There were also no statistically significant differences for the percentage of participants identified as impaired who were referred to the neuropsychological clinic versus other sites where participants were not referred (34.5% versus 31.7%, respectively), $\chi^2(1, N = 235) = 0.16, p = .69$. Other participant demographic information and clinical characteristics can be observed in Table 1.

There were more females than males in each group, and the Impaired Group was significantly older than the Unimpaired Group. In addition, the mean education level for the Impaired Group was 0.8 years lower than for the Unimpaired Group. The score ranges for each group on the MMSE indicate little score overlap. The majority of the Impaired Group participants had scores ranging from 21 to 29 (71.7%). Scores of 21–26 (54.1% of the Impaired Group) are categorized in the “mild cognitive impairment range” according to the *MMSE Clinical Guide* (Folstein et al., 2001). In addition, the majority of participants in the Impaired Group were identified by having scores ≥ 25 on the MMSE (74.36%), with fewer than half falling below cut-scores on the TMT (44.9%) or the WMS-III (30.8%).

Group Differences on GrayMatters® Variables

A MANCOVA, controlling for age and education, found statistically significant differences between the Unimpaired Group and the Impaired Group in the expected direction for the three VDR GrayMatters® variables, Wilks' $\lambda = .65, F(3, 229) = 40.60, p < .01, \eta^2 = .35$. As seen in Table 2, follow-up univariate ANCOVA analyses found that unimpaired participants had significantly better scores than the impaired group in total items recalled correctly on the VDR, the number of false positive errors, and the number of correct responses on distracter tasks.

A second MANCOVA, also controlling for age and education, found a statistically significant difference between the two diagnosed groups for the three DAT variables, Wilks' $\lambda = .94, F(3, 229) = 4.73, p < .01, \eta^2 = .06$. As seen in Table 3, statistically significant differences were found on the follow-up ANCOVA univariate analyses in the expected direction between groups for the three DAT variables as well.

Test-Retest

Tables 3 and 4 reflect the correlation coefficients for individual GrayMatters® scores for the test-retest reliability analysis. VDR reliability coefficients ranged from 0.72 to 0.74 ($p < .001$). The DAT correlation coefficients (using the Spearman Rho because of distributional characteristics of the scores) were lower, ranging from 0.37 to 0.50 ($p < .001$). Comparisons of change in performance from the first testing session to the second testing session indicated no statistically significant differences.

Concurrent Validity

One hundred-sixty participants completed the GrayMatters® procedures, the WMS-III and the TMT. Table 5 reflects the correlations between GrayMatters® scores and the scores on the WMS-III and Trail Making Test, Part B. As noted previously, the nonparametric Spearman Rho was used for the DAT data because of the distributional nature of the DAT data.

The correlations between memory-based measures, the WMS-III six index scores and the three VDR scores, were all statistically significant as hypothesized, with correlation coefficients ranging from $r = -.36$ (visual delayed and VDR false positive errors) to $r = -.61$ (visual immediate and VDR correct). All of the correlations were in the expected direction. Interestingly, the highest correlation was between the VDR distracters scores and the TMT, a measure of executive function. TMT scores have been found to also correlate with measures of working memory (Mahurin et al., 2006). The correlations between the tasks designed to measure executive functioning, the TMT and DAT, were more modest with coefficients ranging from $r = .18$ to $r = -.24$. All of these correlations were statistically significant and in the expected direction.

Discussion

Participants in the Impaired Group had psychometric evidence (using age-related norms) of cognitive dysfunction. None of these participants had previously been evaluated for Dementia or MCI. This group reflected a wide range of impairment in psychometric test performance. The group is intended to represent older individuals in the community who would be considered appropriate candidates for screening. We did not collect follow-up data on most of these participants to determine how many eventually were given a diagnosis of Dementia or MCI. A longitudinal study of GrayMatters® performance and clinical outcomes is in progress.

Scores on the VDR variables (Total Items Recalled Correctly, False Positive Errors, Correct Responses on Distracter Items) and DAT variables (Number of Rules Learned, Total Correct Responses, Number of Perseverative Errors) were significantly better in the Unimpaired Group compared to the Impaired Group. However, the omnibus effect size for the VDR variables was much larger than the DAT variables (.35 versus .06). The overall results are consistent with the findings of Tornatore et al. (2005), and this study supports the use of self-administered computerized assessment procedures as a valid screen for cognitive impairment in older adults. In the Tornatore et al. (2005) study the unimpaired and impaired

groups were defined using a cutoff of 10th percentile in performance on the Logical Memory II subtest of the Wechsler Memory Scale-Revised. The current study included measures of both verbal and nonverbal memory as well as a measure of executive functions to differentiate between impaired and unimpaired participants.

Most participants completed GrayMatters® in 20 min or so, with a few participants taking 30 to 35 min to complete. This study did not include a recording of response times or time to completion of the overall test. Analysis of response times is currently being evaluated to determine whether this information will be clinically useful.

The results of the test–retest reliability and concurrent validity analyses indicate that the VDR scores are stable over time and appear to tap similar memory processes consistent with the WMS-III. The DAT test–retest and validity coefficients, however, were lower than expected. This was not anticipated given that previous research (Benge, 2003) found more robust relationships between the DAT and other executive functions measures with unimpaired older adults. Lower temporal stability estimates can be for numerous reasons, including practice effects, ceiling effects, and the construct measured. There was little evidence of a practice effect during the retest interval as reflected in the lack of statistically significant differences in DAT and VDR scores from the first administration to the second administration one month later. A ceiling effect seems plausible for the DAT Rules variable with a range of only four, but it does not explain the lower coefficients for the other two DAT variables. One potential explanation is that executive functioning measures in general struggle with temporal stability (Dikmen, Heaton, Grant, & Temkin, 1999; Ross et al., 2005). Further research is needed to evaluate the psychometric properties of the DAT. Evaluating the “clinical” reliability and validity may be of particular import given that the purpose is to facilitate dichotomous decision making—impaired or not. The current study demonstrated that impaired and unimpaired participants performed differently on the DAT, but future study needs to evaluate how well it classifies individuals into these categories.

Wild et al. (2008) discussed the importance of available norms, evidence for reliability and validity, and usability as key factors in determining whether and how computerized assessment/screening systems should be used. They also discussed differences among currently available systems in manner of administration (self-administered or administered by an examiner) and potential applications (screening or more broadly based assessment of cognitive abilities). The current study supports the further evaluation of GrayMatters® as a tool for generating reliable and valid scores to detect early changes in cognition.

The GrayMatters® system measures were designed to measure two key domains (memory and executive functions), in contrast to other systems such as Mindstreams (Dwolatzky et al., 2003) which measures a broader range of cognitive domains. The decision to measure a more limited range of cognitive domains was made to enhance the usability of the system in settings where testing time is a critical factor (such as primary health care offices and clinics). The correlational data support the use of the GrayMatters® variables as valid measures of these cognitive domains. The group differences in GrayMatters® performance also supports the assessment of these cognitive domains in differentiating between normal older individuals and older individuals who are experiencing cognitive deficits. As already

mentioned, the VDR findings were more robust relative to the DAT with regard to both differentiating between groups and the correlational data. Future research needs to evaluate if the DAT provides incremental information for diagnostic purposes. For example, further evaluation of GrayMatters® with a sample composed of individuals identified with mild cognitive impairment will help to determine whether the selection and measurement of these two domains will permit adequate sensitivity and specificity for the GrayMatters® system to be a clinically useful decision tool.

The current study had two limitations of note. First, is a portion of the participants in the Impaired Group likely had more than mild cognitive impairment; approximately 28% had an MMSE score below the mild cognitive impairment cut-score. This is a concern given that the purpose of the GrayMatters® system is to identify those with mild cognitive impairment. We believe this concern is tempered by the fact that the remaining 72.1% of the sample had scores more consistent with mild cognitive impairment. Second, was the lack of racial/ethnic and socioeconomic diversity among the study participants. One of the difficulties in clinical research with volunteers from the community is ensuring that adequate ethnic and socioeconomic diversity is represented among the study participants. Efforts are being made to more systematically evaluate the utility of the GrayMatters® system with ethnic minority groups and with a broader range of socioeconomic levels.

Arguments against widespread screening for cognitive dysfunction in older adults have focused on three central issues: the potential fear and anxiety that older individuals may experience when they are assessed for cognitive deficits and possible AD, the affect of screening and early identification on the resources of the health care system, and the accuracy and costs of available screening approaches (Weimer & Sager, 2009). Subsequent research has suggested that the negative response from screened individuals may have been overestimated (Carpenter et al., 2008). The affect on the health care system may actually be very positive rather than negative (Weimer & Sager, 2009). The accuracy and cost-efficiency of screening approaches is being addressed and progress is being made (Dwolatzky et al., 2003; Tornatore et al., 2005; Wild et al., 2008).

The early diagnosis of AD and other Dementias would be aided significantly if a system were in place to alert the primary care physician that a patient is experiencing cognitive decline. It is important at this time to develop approaches that will enable systematic measurement of key cognitive domains in an accurate and cost-effective manner so that appropriate work-up and diagnosis may be initiated. The early identification of symptoms of Dementia is challenging because measurement of memory and executive functions is difficult and time-consuming and because of individual differences in skills at baseline, before developing symptoms of AD or other Dementias. Computerized approaches to Dementia screening are likely to become more widespread (Ashford, et al., 2006; Wild et al., 2008). A self-administered computerized system reduces costs because staff time requirements for test administration and data handling are reduced. For example, in the current study the administration of even a subset of commonly used neuropsychological instruments required 2.5 hr of participant time, whereas the GrayMatters® system required 20 min for most participants to complete, with longer periods (up to 30–35 min) required for some persons. With the economic savings associated with a computerized, self-administered

system, more economical data collection can be achieved for both research and clinical purposes. Large-scale population studies which include evaluation of cognitive dysfunction may be made more efficient and economical by the use of computers for the initial screen. Clinical evaluation of at-risk but asymptomatic individuals may also be done economically if self-administered computerized systems are used. In addition, the economics of the self-administered computerized system makes it more feasible to establish baseline scores in asymptomatic individuals so that changes in scores over time may also be used as an index of the need for further evaluation. Identification of impairment on the basis of change relative to previously measured baseline would be preferable to identification of impairment by comparison to a normative group after a single screening occasion.

Is it time to initiate population-based cognitive screening? Weimer and Sager (2009) have pointed out that there are immediate financial and social benefits as a result of earlier diagnosis but that population-wide screening may be premature until more effective treatments are available. As newer medication treatments are introduced over the next few years, and as other nonmedication treatment and management strategies are demonstrated to be cost-effective, the clinical and behavioral neurosciences will be called on to provide accurate and economical approaches to identifying older individuals who may be experiencing cognitive decline. At that point, the social and financial benefits described by Weimer and Sager (2009) can be maximized.

Acknowledgments

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Albert M, Moss M, Tanzi R, Jones K. Preclinical predication of AD using neuropsychological tests. *Journal of the International Neuropsychological Society*. 2001; 7:631–639. [PubMed: 11459114]
- Alladi S, Arnold R, Mitchell J, Nestor PJ, Hodges JR. Mild cognitive impairment: Applicability of research criteria in a memory clinic and characterization of cognitive profile. *Psychological Medicine*. 2006; 36:507–515. [PubMed: 16426486]
- Alzheimer's Association. 2009 Alzheimer's disease facts and figures. *Alzheimer's and Dementia*. 2009; 5:234–270.
- Archibald S, Kerns K. Identification and description of new tests of executive functioning in children. *Child Neuropsychology*. 1999; 5:115–130.
- Ashendorf L, Jefferson AL, O'Connor MK, Chaisson C, Green RC, Stern RA. Trail making test errors in normal aging, mild cognitive impairment, and Dementia. *Archives of Clinical Neuropsychology*. 2008; 23:129–137. [PubMed: 18178372]
- Ashford JW, Borson S, O'Hara RO, Dash P, Frank L, Robert P, Buschke H. Should older adults be screened for Dementia? *Alzheimer's & Dementia*. 2006; 2:76–85.
- Baddeley A, Baddeley H, Bucks R, Wilcock G. Attentional control in Alzheimer's disease. *Brain*. 2001; 24:1492–1508. [PubMed: 11459742]
- Benge, JF. Unpublished master's thesis. Abilene Christian University; Abilene, Texas: 2003. Evaluation of a computerized alternation task as a measure of executive function in an elderly sample.

- Blackwell AD, Sahakian BJ, Vesey R, Semple JM, Robbins TW, Hodges JR. Detecting Dementia: Novel neuropsychological markers of pre-clinical Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*. 2004; 17:42–48. [PubMed: 14560064]
- Bloom BS, de Povourville N, Straus WL. Cost of illness of Alzheimer's disease: How useful are current estimates? *Gerontologist*. 2003; 43:158–164. [PubMed: 12677073]
- Boustani M, Callahan CM, Unverzagt FW, Austrom MG, Perkins AJ, Fultz BA, Hendrie HC. Implementing a screening and diagnosis program for Dementia in primary care. *Journal of General Internal Medicine*. 2005; 20:572–577. [PubMed: 16050849]
- Boustani M, Peterson B, Hanson L, Harris R, Lohr KN. Screening for Dementia in primary care: A summary of the evidence for the US Preventive Services Task Force. *Annals of Internal Medicine*. 2003; 138:927–942. [PubMed: 12779304]
- Cahn D, Salmon D, Butters N, Wiederholt W, Corey-Bloom J, Edelstein SL, Barrett-Connor E. Dementia of the Alzheimer's type in a population-based sample: Neuropsychological test performance. *Journal of the International Neuropsychological Society*. 1995; 1:252–260. [PubMed: 9375219]
- Carpenter BD, Xiong C, Porensky EK, Lee MM, Brown PJ, Coats M, Morris JC. Reaction to a Dementia diagnosis in individuals with Alzheimer's disease and mild cognitive impairment. *Journal of the American Geriatric Society*. 2008; 56:405–412.
- Chen P, Ratcliff F, Belle S, Cauley D, DeKosky S, Ganguli M. Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology*. 2000; 55:1847–1853. [PubMed: 11134384]
- Collins, B. Unpublished doctoral dissertation. University of Ottawa; Ottawa, Ontario, Canada: 2000. Frontal lobe dysfunction in Alzheimer's disease: Assessment and prognostic significance.
- Dikmen SS, Heaton RK, Grant I, Temkin NR. Test-retest reliability and practice effects of expanded Halstead-Reitan neuropsychological test battery. *Journal of the International Neuropsychological Society*. 1999; 5:346–356. [PubMed: 10349297]
- Duke L, Kaszniak A. Executive control functions in degenerative Dementias: A comparative review. *Neuropsychology Review*. 2000; 10:75–99. [PubMed: 10937917]
- Dwolatzky T, Whitehead V, Doniger GM, Simon ES, Schweiger A, Jaffe D, Chertkow H. Validity of the mindstreams computerized cognitive battery for mild cognitive impairment. *BMC Geriatrics*. 2003; 3:1–12. [PubMed: 12600276]
- The Washington Times. 2005 Jun 16. Fighting Alzheimer's, Saving Tax Dollars.
- Folstein, M.; Folstein, S.; Fanjiang, G. Minimental state examination clinical guide. Lutz, FL: PAR; 2001.
- Ganguli M. The use of screening instruments for the detection of Dementia. *Neuroepidemiology*. 1997; 16:271–280. [PubMed: 9430126]
- Guifford DR, Cummings JL. Evaluating Dementia screening tests: Methodological standards to rate their performance. *Neurology*. 1999; 52:224–227. [PubMed: 9932934]
- Haut MW, Demarest D, Keefover RW, Rankin ED. Controlling for constructional dysfunction with the visual reproduction subtest of the Wechsler Memory Scale-Revised in Alzheimer's disease. *Clinical Neuropsychologist*. 1996; 10:309–312.
- Hunter W. The delayed reaction in animals and children. *Behavior Monographs*. 1913; 2:1–86.
- Jonnson L. Pharmacoeconomics of cholinesterase inhibitors in the treatment of Alzheimer's disease. *Pharmacoeconomics*. 2003; 21:1025–1037. [PubMed: 13129415]
- Lee JE, Park HJ, Song SK, Sohn YH, Lee JD, Lee PH. Neuroanatomic basis of amnesic MCI differs in patients with and without Parkinson disease. *Neurology*. 2010; 75:2009–2016. [PubMed: 21115956]
- Linn R, Wolf P, Bachman D, Knoefel J, Cobb J, Belanger A, D'Agnostino R. The preclinical phase of probable Alzheimer's disease. *Archives of Neurology*. 1995; 52:485–490. [PubMed: 7733843]
- Lorentz WJ, Scanlan JM, Borson S. Brief screening tests for Dementia. *Canadian Journal of Psychiatry*. 2002; 47:723–733.
- Mahurin RK, Velligan DI, Hazleton B, Mark Davis J, Eckert S, Miller AL. Trail making test errors and executive function in schizophrenia and depression. *Clinical Neuropsychology*. 2006; 20:271–288.

- Oscar-Berman M, Zola-Morgan S. Comparative neuropsychology and Korsakoff's syndrome: Spatial and visual reversal learning. *Neuropsychologia*. 1980; 18:499–512. [PubMed: 6449680]
- Oscar-Berman M, Zola-Morgan S, Oberg R, Bonnder R. Comparative neuropsychology and Korsakoff's syndrome: Delayed response, delayed alternation, and DRL performance. *Neuropsychologia*. 1982; 20:187–202. [PubMed: 6211634]
- Partiot A, Verin M, Pillon B, Teixeira-Ferreira C, Agid Y, Dubois B. Delayed response tasks in basal ganglia lesions in man: Further evidence for a striato-frontal cooperation in behavioral adaptation. *Neuropsychologia*. 1996; 34:709–721. [PubMed: 8783222]
- Perry R, Hodges J. Attention and executive deficits in Alzheimer's disease: A critical review. *Brain*. 1999; 122:383–404. [PubMed: 10094249]
- Reitan, RM.; Wolfson, D. The Halstead-Reitan neuropsychological test battery. Theory and clinical interpretation. Tuscany, Italy: Neuropsychology Press; 1985.
- Ross TP, Weinberg M, Furr AE, Carter SE, Evans-Blake L, Parham S. The temporal stability of cluster and switch scores using a modified COWAT procedure. *Archives of Clinical Neuropsychology*. 2005; 20:983–996. [PubMed: 15990269]
- Saunders NL, Summers MJ. Longitudinal deficits to attention, executive, and working memory in subtypes of mild cognitive impairment. *Neuropsychology*. 2011; 25:237–248. [PubMed: 21381828]
- Scharre DW, Chang SI, Murden RA, Lamb J, Beversdorf DQ, Kataki M, Bornstein RA. Self-administered gerocognitive exam (SAGE): A brief cognitive assessment instrument for mild cognitive impairment (MCE) and early Dementia. *Alzheimer Disease and Associated Disorders*. 2010; 24:64–71. [PubMed: 20220323]
- Sgaramella T, Borgo F, Mondini S, Pasini M, Toso V, Semenza C. Executive deficits appearing in the initial stage of Alzheimer's disease. *Brain and Cognition*. 2001; 46:264–268. [PubMed: 11527345]
- Snitz BE, Morrow LA, Rodriguez EG, Huber KA, Saxton JA. Subjective memory complaints and concurrent memory performance in older patients of primary care providers. *Journal of the International Neuropsychological Society*. 2008; 14:1004–1013. [PubMed: 18954480]
- Tierney MC, Lerner MA. Computerized cognitive assessment in primary care to identify patients with suspected cognitive impairment. *Journal of Alzheimer's Disease*. 2010; 20:823–832.
- Tornatore JB, Hill E, Laboff JA, McGann ME. Self-administered screening for mild cognitive impairment: Initial validation of a computerized test battery. *Journal of Neuropsychiatry and Clinical Neuroscience*. 2005; 17:98–105.
- Wechsler, D. WMS-III administration and scoring manual. San Antonio, TX: Psychological Corporation; 1997.
- Weimer DL, Sager MA. Early identification and treatment of Alzheimer's disease: Social and fiscal outcomes. *Alzheimer's and Dementia*. 2009; 5:215–226.
- Wild K, Howieson D, Webbe F, Seelye A, Kaye J. Status of computerized cognitive testing in aging: A systematic review. *Alzheimer's & Dementia*. 2008; 4:428–437.

Biographies

Samuel D. Brinkman, PhD is a neuropsychologist and owner of the Neuropsychology Clinic in Abilene, TX. He is also president of Dementia Screening, Inc. Research interests include aging and the dementias, traumatic brain injury, and recovery from brain disorders.

Robert J. Reese, PhD, is associate professor of counseling psychology at the University of Kentucky in the Department of Educational, School, and Counseling Psychology. His research interests include psychotherapy process and outcome and the utilization of technology to provide psychological services to underserved populations.

Larry A. Norsworthy, PhD, is professor of psychology at Abilene Christian University in the Department of Psychology. His research interests include psychotherapy process and outcome and chronic pain.

Donna K. Dellaria, MA, is a Licensed Professional Counselor in Seguin, Texas. Her research interests include personality types and cognitive development across the life span.

Jacob W. Kinkade, MA, is a Licensed Professional Counselor in Abilene, TX and a PhD student at Walden University.

Jared Benge, PhD, ABPP is a clinical neuropsychologist for the Jack C. Montgomery VA Medical Center in Muskogee, OK. His research interests include psychometrics and outcomes from traumatic brain injury.

Kimberly Brown, RN is a Clinical Research Coordinator at the University of North Texas Health Science Center. Research activities involve coordination of translational research studies related to the aging process.

Anna Ratka, PhD, PharmD is a professor at the College of Pharmacy Texas A&M Health Science Center. Her research expertise is in neuropharmacology of pain, effects of age and neurodegeneration on pain, and effect of menopause on hormones on women cognitive health.

James W. Simpkins, PhD conducts research on brain aging, Alzheimer's disease and stroke and the discovery of drugs for these age-related brain disorders. He has authored 348 scientific papers on this subject, and has drugs in clinical trials for effectiveness in a variety of indications. He is the Director of the Institute for Aging and Alzheimer's Disease Research at the University of North Texas Health Science Center.

Table 1

Demographic and Clinical Characteristics of Groups.

Demographic	Entire group (<i>N</i> = 235)	Unimpaired participants (<i>n</i> = 157)	Impaired participants (<i>n</i> = 78)	<i>p</i> -value
Age	74.77 (8.6)	72.20 (7.6)	79.94 (8.3)	.001 ^a
Education (yrs.)	13.39 (2.7)	13.66 (2.8)	12.85 (2.4)	.05 ^a
Gender				
Male	30%	28%	33%	.702 ^b
Female	70%	72%	67%	
MMSE Score	26.73 (4.7)	29.11 (0.9)	21.95 (5.6)	.001 ^a
Range	5–30	27–30	5–29	

Note. Numbers outside of the parentheses reflect the mean scores and the numbers and the numbers within the parentheses reflect the standard deviation scores. MMSE = mini mental state exam.

^a *t*-test for independent samples, two-tailed.

^b Chi-square.

Table 2

Comparisons for GrayMatters® Variables ($N = 235$).

Variable	Full group ($N = 235$)	Unimpaired ($n = 157$)	Impaired ($n = 78$)	F	η^2
VDR recall	18.41 (3.0)	19.61 (2.2)	16.00 (3.0)	54.52 ***	.42
VDR false positive errors	3.52 (2.4)	2.87 (1.9)	4.83 (2.7)	20.90 ***	.21
VDR distracters correct	9.12 (2.1)	10.01 (1.2)	7.35 (2.4)	60.51 ***	.44
DAT rules	3 (0–4)	3 (0–4)	2 (0–4)	12.51 ***	.14
DAT total correct	74 (30–96)	78 (33–96)	70 (30–94)	14.37 ***	.15
DAT perseveration errors	10 (0–56)	8 (0–48)	11 (0–56)	9.06 ***	.10

Note. VDR = visual delayed recognition; DAT = delayed alternation task. For VDR variables, numbers outside of the parentheses reflect the mean scores and the numbers and numbers within the parentheses reflect the standard deviation scores. For DAT variables, numbers outside of the parentheses reflect the median scores and the numbers within the parentheses reflect the range.

 $p < .001$.

Table 3

Test–Retest Reliability and Differences in VDR Scores After 30-Day Interval ($N = 90$).

	First test session	Second test session	Pearson correlation coefficient
VDR recall	18.83 (2.8)	19.12 (3.3) ^a	0.72***
VDR false positive errors	3.17 (2.4)	3.12 (2.6) ^a	0.73***
VDR distracter correct	9.12 (2.4)	9.23 (2.1) ^a	0.74***

Note. VDR = visual delayed recognition. Numbers outside of the parentheses reflect the mean scores and the numbers within the parentheses reflect the standard deviation scores.

^a Changes in scores from first session to second session not statistically significant by within-groups *t*-test.

 $p < .001$.

Table 4

Test–Retest Reliability and Differences in DAT Scores After 30-Day Interval ($N = 90$).

	First test session	Second test session	Spearman correlation coefficient
Rules correct	2.83 (1.0)	2.74 (0.9) ^a	0.37***
DAT total correct	74.19 (10.9)	75.23 (9.4) ^a	0.46***
DAT perseverative errors	10.31 (7.3)	9.28 (4.8) ^a	0.50***

Note. DAT = delayed alternation task. Numbers outside of the parentheses reflect the mean scores and the numbers within the parentheses reflect the standard deviation scores.

^a Changes in scores from first session to second session were not statistically significant using a wilcoxon signed ranks test.

 $p < .001$.

Table 5Correlations of GrayMatters® Scores With WMS-III Subtest Scores and TMT Scores ($N = 160$).

	Auditory immediate	Auditory delayed	Auditory recognition	Visual immediate	Visual delayed	TMT time
VDR ^a correct	.61***	.57***	.54***	.58***	.54***	-.49***
VDR ^a distractors	.49***	.45***	.48***	.41***	.37***	-.62***
VDR ^a false positive	-.55***	-.51***	-.51***	-.41***	-.36***	.39***
DAT ^b rules	.18*	.16*	.12	.09	.09	-.21**
DAT ^b total correct	.21**	.21**	.18*	.15	.16*	-.24**
DAT ^b perseverative	-.19*	-.21**	-.19*	-.15	-.15	.18*

Note. VDR = visual delayed recognition; DAT = delayed alternation task.

^a Pearson correlation coefficients are reported.^b Spearman Rho correlation coefficients are reported.* $p < .05$,** $p < .01$,*** $p < .001$.