

Review Article

Status of computerized cognitive testing in aging: A systematic review

Katherine Wild^{a,*}, Diane Howieson^a, Frank Webbe^b, Adriana Seelye^a, Jeffrey Kaye^a

^aLayton Aging and Alzheimer Center, Oregon Health and Science University, Portland, OR, USA

^bFlorida Institute of Technology, Melbourne, FL, USA

Abstract

Background: Early detection of cognitive decline in the elderly has become of heightened importance in parallel with the recent advances in therapeutics. Computerized assessment might be uniquely suited to early detection of changes in cognition in the elderly. We present here a systematic review of the status of computer-based cognitive testing, focusing on detection of cognitive decline in the aging population.

Methods: All studies purporting to assess or detect age-related changes in cognition or early dementia/mild cognitive impairment by means of computerized testing were included. Each test battery was rated on availability of normative data, level of evidence for test validity and reliability, comprehensiveness, and usability. All published studies relevant to a particular computerized test were read by a minimum of two reviewers, who completed rating forms containing the abovementioned criteria.

Results: Of the 18 test batteries identified from the initial search, 11 were appropriate to cognitive testing in the elderly and were subjected to systematic review. Of those 11, five were either developed specifically for application with the elderly or have been used extensively with that population. Even within the computerized testing genre, great variability existed in manner of administration, ranging from fully examiner-administered to fully self-administered. All tests had at least minimal reliability and validity data, commonly reported in peer-reviewed articles. However, level of rigor of validity testing varied widely.

Conclusion: All test batteries exhibited some of the strengths of computerized cognitive testing: standardization of administration and stimulus presentation, accurate measures of response latencies, automated comparison in real time with an individual's prior performance as well as with age-related norms, and efficiencies of staffing and cost. Some, such as the Mild Cognitive Impairment Screen, adapted complicated scoring algorithms to enhance the information gathered from already existing tests. Others, such as CogState, used unique interfaces and subtests. We found that although basic indices of psychometric properties were typically addressed, sufficient variability exists that currently available computerized test batteries must be judged on a case-by-case basis.

© 2008 The Alzheimer's Association. All rights reserved.

Keywords: Computerized cognitive assessment; Computerized testing; Early detection; Systematic review; Dementia; Psychometrics

1. Introduction

With the aging of the population, the rapidly growing incidence of dementia has become a major public health concern. The development of new therapies has been a promising response to this health care crisis. However, to be optimally

effective, treatments must be started early in the disease process. Thus, detection of cognitive decline in the elderly has become of heightened importance in parallel with the recent advances in therapeutics. In addition, once therapies are established as effective interventions, measures for following their effectiveness will be needed, especially if there is potential toxicity associated with a particular treatment.

Computerized administration of clinical instruments is not an entirely new phenomenon. The first personal com-

*Corresponding author. Tel.: 503-494-6975; Fax: 503-494-7499.
E-mail address: wildk@ohsu.edu

puters were introduced into wide use in the 1970s. Rapid adoption of computer-based testing paralleled this development. By the 1980s the research literature was replete with considerations of the inherent advantages and limitations of automated assessment of a myriad of clinical domains. In particular, the application of computers to the evaluation of cognition has been widely studied. This body of research has generally fallen into one of two categories: (1) the translation of existing standardized tests to computerized administration and (2) the development of new computer tests and batteries for the assessment of cognitive function. Somewhere between these two categories are approaches that have adapted an existing test in a new way by using computer administration. This review will focus on those tests and batteries that have been applied to or developed specifically for the detection of cognitive changes in the elderly.

Computerized assessment might be uniquely suited to early detection of changes in cognition in the elderly. Among the multiple advantages that have been cited, computer tests can cover a wider range of ability, minimize floor and ceiling effects, are given in a standardized format, and precisely record accuracy and speed of response with a level of sensitivity not possible in standard administrations. Such characteristics can be critical both in early detection and in extending the range of a test to be sensitive both to mild cognitive impairment (MCI) and also to the more pronounced changes occurring in the early stages of dementia. In comparison with traditional neuropsychological assessment instruments, computerized tests might also represent a potential cost savings, not only with regard to materials and supplies but also in the time required of the test administrator. Moreover, the nature of the computerized instruments might allow administration by health care associates other than neuropsychologists, as long as the critical activities of interpretation and diagnosis are performed by the appropriate professional. A full discussion of the ethical issues that confront the developers and users of computerized neuropsychological assessment instruments is beyond the scope of this article, but several recent reviews have addressed this topic [1,2]. Recognition of the utility of computer-administered clinical assessments has coincided with recent and rapid advances in personal computer technology, producing numerous test batteries. Furthermore, with the widespread availability of the Internet, computer-based cognitive assessment provides the potential for large-scale screening of populations for cognitive function.

In the initial excitement of this new application of technology, however, some basic aspects of test development might have been sacrificed. One of the more persistent criticisms of computerized test batteries is the general lack of adequately established psychometric standards [3]. Other concerns include failure to demonstrate equivalence between the examinee's experience of computer versus traditional test administration, limited—and for the elderly, per-

haps unfamiliar—response modality, and poorly designed computer-person interface. The goal of this review was not to present a critique of computer versus standard cognitive testing, although some such comparisons were unavoidable, but to examine the more widely used and researched computer batteries across a uniform set of test criteria. To this end, we present here a systematic review of the status of computer-based cognitive testing, focusing on detection of cognitive decline in the aging population.

2. Methods

Because of the heterogeneity across selected studies and test batteries, a meta-analytic approach was not possible. The rigorous methodology used by that technique was not suited to the current published research in this relatively new field, given the variety in types of studies, subjects, and data analyses. Guided by the Cochrane Collaboration's [4] published recommendations, the present report represents findings from a systematic review of currently available computerized test batteries for the detection of cognitive change in the elderly.

2.1. Search strategies

A search was performed at the end of 2006 and again in early 2007 of the databases of PubMed, PsycInfo, and Cochrane. The following Medical Subject Headings (MeSH) and PsycInfo headings were used: technology assessment (biomedical), assessment (technology), biomedical technology assessment, and computerized cognitive assessment. Keywords from retrieved articles included computerized testing, computer, multimedia, and computerized battery and were used in a second search. From the literature collected on that basis, reference lists were examined individually for citations of additional relevant studies. This search resulted in 79 citations.

2.2. Inclusion and exclusion criteria

All studies purporting to assess or detect age-related changes in cognition or early dementia/MCI by means of computerized testing were included. Studies in which appropriate batteries were administered to different populations (eg, subjects with multiple sclerosis) or for different applications (eg, driving safety evaluation) were excluded. Tests that were developed for the target population but were not available in English were excluded, as were tests for which separate norms for the elderly were not available. Computerized systems that were reported to be too difficult for cognitively impaired elderly to navigate were also excluded from this review. Table 1 summarizes those batteries that were identified by the search strategies described above but were excluded from review on closer examination.

2.3. Rating methodology

Each test battery was rated on availability of normative data and level of evidence for test validity and reliability.

Table 1
Computerized tests excluded from comprehensive review

Test Name	First Author, Year	Reason for Exclusion
Automated Cognitive Test Cognometer	Stollery B, 1996 Polich J, 2001	Developed for use in neurotoxicology Mean age, 44 y; no separate data for elderly
Computerized Neurocognitive Scan Examen Cognitif par Ordinateur (ECO)	Gur RC, 2001 Ritchie K, 1993	Mean age, <30 y; no separate data for elderly French language only.
Hasegawa Dementia Scale	Inoue M, 2000	Inadequate data; available in Japan
Integneuro	Paul RH, 2005	Not specific to elderly.
Neurobehavioral Evaluation System (NES)	Letz R, 1996; White RF, 2003	Developed as neurotoxicology screen: not specific to elderly

The test batteries were also evaluated in terms of comprehensiveness, usability, cost, and availability. All studies relevant to a particular computerized test were read by two reviewers, who completed rating forms containing the abovementioned criteria. All ratings for a particular test were then combined to yield an overall rating. For example, if one study dealt comprehensively with a test's reliability and another with its construct validity, that test would be awarded maximum points on both parameters. (See the Appendix for the rating template).

3. Results

Eleven test batteries were subject to systematic review as described above. A summary of some basic parameters is presented in Table 2. For many key test variables such as length of administration, alternate form availability, and means of generation of results, information was inconsistently available. These data are therefore presented in the text whenever obtainable rather than in tabular form, because direct comparisons were not possible.

Table 3 presents overall ratings of each test battery. Unfortunately, information regarding cost and availability was so infrequently reported that these variables were removed as a basis of comparison.

3.1. Automated Neuropsychological Assessment Metrics

Originally developed for use by the Department of Defense, this battery has been applied to several clinical populations, including cognitively impaired elderly. A subset of six of the 30 Automated Neuropsychological Assessment Metrics (ANAM) tests form the ANAM Dementia Screening Battery: simple and choice reaction times, matching to sample, continuous performance test, a Sternberg six-letter memory task, and spatial discrimination. The test is administered with an examiner present to clarify instructions or use of the mouse as needed. No information on duration of this battery was presented. Correlations between traditional neuropsychological tests and ANAM subtests of allegedly similar cognitive domains are reported by Kabat et al [5]. The study describes a principal components analysis that yielded three factors: processing speed/efficiency, retention/

Table 2
Test battery descriptions

Test	Age Range (y)	Largest Sample	Administration	Domains*
ANAM	22–77	191	Mouse/keyboard; self-administered	Memory, attention, psychomotor speed, language, reaction time
CANS-MCI	51–93	310	Touch screen; self-administered	Memory, language, executive function
CANTAB	8–80	771	Touch screen/keyboard; technician administered	Working memory, attention, visuospatial memory
CNS Vital Signs	7–90	1069	Keyboard; self-administered	Memory, psychomotor speed, processing speed, cognitive flexibility, sustained attention
CNTB	21–87	209	Keyboard; technician administered	Language, information processing, motor speed, attention, spatial, memory
COGDRAS-D	67–103	190	Yes/no button; technician administered	Memory, attention, reaction time†
CogState	18–40; 46–82	113	Keyboard; self-administered	Working memory, executive function, attention, reaction time
CSI	18–89	284	Keyboard; self-administered	Memory, attention, response speed, processing speed
MCIS	>65	215	Technician records responses, or via telephone	Memory, executive function, language
MicroCog	18–89	810	Keyboard/# pad; self-administered	Memory, attention, reaction time, spatial ability, reasoning/calculation
Mindstreams	>50	213	Mouse/# pad; technician administered	Memory, executive function, visuospatial, verbal fluency, attention, motor skills, information processing

* Cognitive domains as identified by authors; subtests intended to measure those domains vary.

† Domains assessed by COGDRAS-D were not identified; eight subtests listed appear to assess these domains.

Table 3
Ratings of included test batteries*

Test Battery	Subtests	Normative Data	Reliability	Validity	Factor Analysis	Admin/Interface
ANAM	2	2	1	3	3	2
CANS-MCI	2	3	3	3	3	3
CANTAB	3	3	2	3	3	2
CNS Vital Signs	3	3	2	3	1	2
CNTB	3	1	3	3	1	2
COGDRAS-D	2	2	2	3	3	2
CogState	3	2	2	2	1	2
CSI	2	2	3	3	3	2
MCIS	1	3	2	2	3	2
MicroCog	3	3	2	3	1	3
Mindstreams	3	2	1	2	1	2

* A score of 1 in general indicates failure to address that component; a rating of 2 represents partial consideration and/or presentation of the variable; a rating of 3 indicates that comprehensive information regarding that element of test construction has been made available.

memory, and working memory. Notably absent from this dementia screening battery are tests of language and delayed memory. In a comparison of a small sample of Alzheimer's disease patients and age-matched controls, Levinson et al [6] reported that a discriminant function analysis of a subset of ANAM scores correctly classified 100% of subjects. However, they also reported that most patients and some controls exhibited some confusion with procedures of the ANAM and suggested modifications on the basis of their findings. A recent supplement to the *Archives of Clinical Neuropsychology* presents a comprehensive series of articles that summarize the test development and psychometric properties of this battery [7–10]. Reliability data are absent from these reports.

3.2. Computer-Administered Neuropsychological Screen for Mild Cognitive Impairment

Developed as a screening instrument for detection of MCI, tests of the Computer-Administered Neuropsychological Screen for Mild Cognitive Impairment (CANS-MCI) are intended to assess language (picture naming), memory (immediate and delayed recall, recognition), and executive function. The executive function tests assess mental control and spatial ability. The battery is described as fully self-operated, although it is designed for use as a screening tool to be used in a clinical setting where an office assistant needs to set the patient up to get started (a process that is said to take minutes; personal communication with CANS-MCI representative, 2007); tests are administered by computer by means of proprietary hardware (a computer with a touch screen and speakers) and are reported to take about 30 minutes to complete. Scoring and report generation is web-based. Information regarding cost and contact information is available via a website (www.mildcognitiveimpairments.com). In the single published study located by the search strategies described above, Tornatore et al [11] administered the CANS-MCI and a battery of standard neuropsychological tests to 310 community-residing elderly. They

reported measures of internal consistency and test-retest reliability, correlations with parallel conventional tests, and differences on the CANS-MCI between memory-impaired and memory-intact groups. A factor analysis supported their model of three factors: memory, language/spatial fluency, and executive function/mental control. Longitudinal studies are reported to be underway to evaluate the battery's sensitivity to change.

3.3. Cambridge Neuropsychological Test Automated Battery

This battery focuses on three cognitive domains: working memory and planning (spatial span, spatial working memory, spatial planning), attention (set shifting, reaction time, visual search), and visuospatial memory (pattern and spatial recognition, delayed matching to sample, paired associate learning). Responses are via touch screen, and the battery is described as largely independent of verbal instruction. Currently this is the most widely published battery, although most reports are based on a small subset of the 13 tests. A principal components analysis based on more than 770 normal controls identified two factors: general learning and memory and speed of response [12]. In one of the earliest reports, Sahakian and Owen [13] identified the paired associates, delayed matching to sample, and attentional set shifting subtests of the Cambridge Neuropsychological Test Automated Battery (CANTAB) as particularly sensitive to differences between healthy controls and early stage Alzheimer's disease and Parkinson's disease patients. The early review by Fray et al [14] supported the application of the CANTAB to assessment of other neurodegenerative disorders. Subsequent studies have reported test-retest reliabilities [15], normative data based on a large sample of healthy elderly subjects [12,16], and early detection of memory deficits [17–19]. These studies limited their investigation to two or three of the most frequently researched subtests.

3.4. CNS Vital Signs

This test battery includes seven tests covering five domains: memory (verbal and visual recognition), psychomotor speed (finger tapping, symbol digit coding), reaction time, cognitive flexibility (shifting attention, Stroop paradigm), and complex attention (continuous performance, shifting attention, Stroop). Tests of memory are administered in a recognition format; no measures of free recall are included. The tests are self-administered and take approximately 30 minutes to complete. Responses are via computer keyboard. Recommended uses are for screening and for serial assessments rather than as a diagnostic tool. Tests of memory, processing speed, and cognitive flexibility have been shown to discriminate between normal control and MCI subjects and between MCI and patients with mild dementia [20]. Gualtieri and Johnson [21] present normative data based on more than 1,000 subjects across the lifespan, with norms broken down by decade; however, they acknowledge the need for expanded norms at the older age ranges. Test-retest reliabilities, correlations with conventional tests, and comparisons across diagnostic groups are summarized. An appendix provides detailed descriptions of each subtest. The authors state that the tests are “familiar and well-established,” but some are significantly modified from the standard format and administration (eg, verbal memory test is said to be adapted from the Rey AVLT, but it presents 15 words for a single trial followed by immediate and delayed recognition, with no free recall trial).

3.5. Computerized Neuropsychological Test Battery

This battery consists of 11 subtests assessing motor speed (finger tapping), information processing (simple reaction time, choice reaction time), attention, verbal and spatial memory (word-list learning and recall, paired associate learning, visual memory, visual matching delayed recall), language (20-item Boston Naming), and spatial abilities (visual matching). The Computerized Neuropsychological Test Battery (CNTB) represents one of the earlier efforts in computer testing of cognitive function (22). Although it is computerized in terms of stimulus presentation and reaction time recording, it is fully administered by a technician and thus is not a self-administered, automated test battery. Test responses require use of a single key, pointing, or spoken responses that are entered by the technician. Developed as an alternative to the Alzheimer's Disease Assessment Scale (ADAS), this battery has been used in clinical trials for the treatment of Alzheimer's disease [23,24]. Although highly correlated with scores on the ADAS, the CNTB was reported to be more sensitive to treatment effects in mildly impaired Alzheimer's disease patients than the ADAS.

3.6. Cognitive Drug Research Computerized Assessment System

This battery was not developed specifically for detection of cognitive decline among the elderly but to measure drug effects, both positive and negative, on cognition in a variety of patient populations. It has, however, been adapted for use with dementia patients (COGDRAS-D). The battery contains eight subtests: immediate word recognition, simple and choice reaction times, memory scanning, digit vigilance, delayed word recognition, delayed picture recognition, and delayed face recognition. All stimuli for this battery are presented on the computer screen. The examiner provides instructions and initiates each task; subjects respond via two buttons (Yes/No), which was found in pilot work to be usable by impaired subjects. No information was found on length of time to administer the entire battery. Test-retest reliabilities and correlations between COGDRAS-D subtests and other clinical assessments are reported by Simpson et al [25]. Mohr et al [26] compared the ability of several clinical trial and standard test batteries to distinguish between normal subjects and mildly demented Alzheimer's disease and Huntington's disease patients. They found that although all batteries (ADAS, Mattis Dementia Rating Scale, Mini-Mental State Examination [MMSE], Wechsler Memory Scale-Revised, and the Repeatable Battery for the Assessment of Dementia) yielded significant differences between the healthy and patient groups, the COGDRAS was the most sensitive in differentiating between the Alzheimer's disease and Huntington's disease groups. Tests of attention from the COGDRAS-D were found to discriminate between patients with Alzheimer's disease and those with Lewy body dementia [27]. However, DeLepeliere et al [28] found that selected subtests of the COGDRAS-D added little diagnostic value to basic clinical tests in a general practice.

3.7. CogState

Subtests include measures of simple, choice, and complex reaction times, continuous monitoring, working memory, matching, incidental learning, and associative learning. All subtests are based on playing card formats, with little reliance on or assessment of verbal abilities. Written instructions are presented on the screen. Responses are made on a computer keyboard represented on the screen, with responses by using the “k” key for yes and “d” key for no. The battery requires 15 to 20 minutes to complete. No data were available on correlations between these tests and standard cognitive assessments. A website presents relevant ordering/cost information (www.cogstate.com). This battery was developed for repeat testing; the authors have therefore published extensive results concerning practice effects [29,30]. Darby et al [31] found that although subjects with “mild MCI” were indistinguishable from healthy controls on initial testing, by the third and fourth administra-

tion
mu
pa:
co:
sul
fol
Es
roj
tes
en-
ad
the

3.8

ce:
ide
sar
ser
av
ba
ba
co
rec
vic
av.
[31
va
su
au
scr
CS
nit
thi
fu
im
ple
wh
im

3.9

ba
co:
Sc
fu
co
sig
tin
ev.
ma
rec
dis
im

tions within a 3-hour span, MCI subjects demonstrated much smaller practice effects than their healthy counterparts. In a comparison of MCI and healthy controls, the continuous learning task of the CogState battery detected subtle changes in memory in subjects with MCI at 12-month follow-up that were not detected by the Consortium to Establish a Registry for Alzheimer Disease (CERAD) neuropsychological test battery or the paired associate learning test of the CANTAB [32]. Cargin et al [33] reported differences on subtests of the CogState between healthy older adults who were separated into two groups on the basis of their CERAD Word List Delayed Recall performance.

3.8. Cognitive Stability Index

Four factors (memory, attention, response speed, processing speed) that are based on 10 subtests have been identified in an analysis of the performance of a normative sample of 18- to 89-year-olds [34]. Instructions are presented on the computer screen, with a test administrator available for clarification. All stimuli are presented nonverbally; responses are made via a restricted set of keys. The battery is estimated to take between 25 to 35 minutes to complete. This is a web-based test with automated online record keeping; therefore, immediate comparison with previous performance is available. Ordering information is available online at www.headminder.com. Erlanger et al [34] also present data on test-retest reliability, concurrent validation against other existing neuropsychological measures, and comparisons of neurology outpatient groups. The authors recommend the Cognitive Stability Index (CSI) for screening and monitoring of change. A shorter variant of the CSI aimed at community screening for dementia, the Cognitive Screening Test (CST), reduces the CSI subtests to three: keyboard skills, learning and memory, and executive function. Lichtenberg et al [35] reported the CST to have impressive concordance with consensus diagnoses in a sample of 102 patients who presented at a geriatric clinic and who were categorized as demented, MCI, or no cognitive impairment.

3.9. MCI Screen

Based on the CERAD word-list learning task, this brief battery is examiner-administered or via telephone (which accounts for the absence of any visuospatial subtests). The MCI Screen (MCIS) is described as assessing memory, executive function, and language. This 10-minute test is essentially a computerized version of the CERAD word-list task, with significant modification. A 10-word list is presented three times, with immediate recall after each presentation; a self-evaluation of memory and a distraction task of triadic animal comparisons are inserted before delayed free and cued recall trials. Finally, free recall of animal names from the distracter task completes the battery. Results are generated immediately on completion of the test, with a primary out-

come of “normal” or “impaired.” Test-retest reliability is based on paper and pencil versus the online MCIS, administered up to 6 months apart. A comparison with the MMSE and Clock Drawing Test found the MCIS to be more sensitive to early detection of cognitive impairment in a primary care setting [36]. A table included in the appendix of that publication provides comparisons of the MCIS with other standard and computerized tests in terms of accuracy in discriminating normal aging from MCI. A unique correspondence analysis (not unlike principal components) algorithm incorporates the scores for each word in the three immediate and one delayed recall administrations. Published research with the standard, noncomputerized form of the test showed that this methodology improved diagnostic sensitivity of detecting MCI by 12% over use of the CERAD delayed recall score and by 9% over the aggregate of the three CERAD learning trials [37].

3.10. MicroCog

Formerly known as the Assessment of Cognitive Skills (ACS), this battery is now marketed by Psych Corp as MicroCog. Development was originally funded by an insurance company to screen older physicians for cognitive impairment (and malpractice risk). The battery consists of a standard form (18 subtests) and short form (12 subtests). Domains include attention/mental control, memory, reasoning/calculation, spatial processing, and reaction time; there are multiple subtests for most domains. This battery is self-administered, with a restricted set of keyboard responses in a multiple-choice format; instructions are presented on the computer screen. Completion of all subtests is reported to take approximately 1 hour for cognitively intact individuals. The authors acknowledge that impaired subjects might take much longer than the estimated 60 minutes; even limited use of the keyboard was reported to cause anxiety and frustration in that group. Green et al [38] found the ACS to differentiate between healthy controls and patients with MCI. In a review of the psychometric properties of the MicroCog, Elwood [39] acknowledged that validation against traditional neuropsychological tests has yielded modest results. Since then, MicroCog’s short form General Cognitive Functioning score and Full Scale IQ from the Wechsler Adult Intelligence Scale-III were reported to be significantly correlated [40]. More recently, Helmes and Miller [41] found weak correlations between measures of the MicroCog and corresponding measures from the Wechsler Memory Scale-III.

3.11. Mindstreams (Neurotrax)

This battery consists of nine subtests: verbal memory, nonverbal memory, Go-NoGo response inhibition, Stroop interference, problem solving, visual spatial imagery, verbal rhyming, verbal naming, staged information processing speed, finger tapping, and visuomotor planning. Several

subtests represent adaptations or analogues of familiar paper and pencil tests such as the Benton Visual Retention Test, Stroop, and the Test of Nonverbal Intelligence. It was designed for use with the elderly in the detection of MCI. Responses are via a mouse and the number pad of the computer keyboard; the entire battery is reported to take 45 to 60 minutes to complete. An abbreviated version has been found to be equally effective in discriminating between persons with mild dementia, MCI, and healthy controls [42]. The full battery has been shown to discriminate between MCI and healthy elderly at a rate comparable to traditional neuropsychological tests of similar domains [43,44]. Doniger et al [45] found that the ability of the Mindstreams battery to detect early cognitive impairment was unaffected by the presence of depressive symptoms in a cohort of normal, MCI, and Alzheimer's disease subjects. Additional references and information are presented at www.neurotrax.com.

4. Discussion

In our systematic search, we identified 11 computer-based test batteries that either were developed to screen for cognitive decline in the elderly or have been applied to that function. Of the 11, five were developed specifically as cognitive impairment screens, whereas the remainder were developed for different purposes and have been adapted or simply co-opted to that end. In all cases, published research describing psychometric properties of these relatively new additions to the field of neuropsychological assessment was identified.

Although most tests reviewed here presented what we judged to be sufficient data to demonstrate validity, other standard measures of test quality were less comprehensively addressed. In just more than half of the batteries, normative data for elderly subjects were rated as less than adequate as a result of either small sample size or lack of data specific to older adults in a larger sample. Reliability data were typically presented in some form, although only three test batteries met our highest rating achieved by describing more than one type of reliability. Factor analytic-type data were reported for six batteries. It is interesting to note that the number of publications devoted to a particular battery had little bearing on its overall ratings. The CANS-MCI, for which only one article was located [11], presented norms for 310 community-dwelling older adults. Test-retest and internal consistency reliability, concurrent and discriminant validity, and confirmatory factor analysis data were also reported.

Whether included in the test development phase or as post hoc analyses, basic indices of psychometric properties are essential to the widespread acceptance of new cognitive test batteries. Schlegel and Gilliland [3] have outlined the necessary elements of quality assurance assessments for computer-based batteries. They caution against the acceptance of computerized adaptations of paper and pencil tests

purely on the basis of face validity. Others have also warned that equivalence across these media cannot be assumed [46–48]. At a minimum, differences in communication of instructions, stimulus presentation, and response format might yield significant differences in test performance, particularly in an older population. For example, computer-based memory tests mostly rely on memory recognition rather than memory free recall, a measure sensitive to memory decline with both aging and MCI. Finally, differences in computer experience among elders as an intervening variable in performance have been largely ignored. Raymond et al [49] did find significant practice effects in administration of the MicroCog 2 weeks apart, which they hypothesized to be at least partly a result of increase in confidence with use of the computer. They suggest that more comprehensive training and familiarity with this relatively novel format before assessment might yield a more reliable measure of change.

Nevertheless, multiple advantages of computer-based assessments over traditional instruments have been identified [50,51]. A frequently cited benefit of computerized testing has been the cost savings and scheduling flexibility offered by the reduced need for administration by trained personnel. We found that few of these batteries are fully self-administered. Indeed, the tests reviewed here range widely in the amount of interaction required of an examiner. Some, like the CANS-MCI and MicroCog, are reported to be entirely self-administered, with instructions via computer screen or speakers. At the other end of the spectrum, the CNTB is described as a computer-assisted battery in which the technician presents the stimuli and records the responses. Most fall somewhere in between, with a technician present during the examination to provide instruction and/or initiate each task in the battery. This variability suggests that the benefits of standardized administration that have been attributed to computerized tests should be examined on a case-by-case basis. Moreover, the use of computerized instruments when dementia is suspected and when the participant is elderly presents critical issues relevant to the cognitive strengths and weakness of the participant, the nature of the person-computer interface, and the impact of technology generally on aged people. The presence of an intermediary who can clarify instructions so as to elicit valid performance or who can halt an assessment process that is invalid might be necessary for ethical reasons.

Another potential advantage of computerized test batteries over traditional paper and pencil assessments is their flexibility in terms of immediate adjustment to performance levels. Many batteries have the capability of automatically altering test order, presentation rate, and level of difficulty in response to test performance. This built-in responsiveness, however, does not always allow the examiner to “test the limits” of ability in pursuit of valuable clinical information. Furthermore, item response theory would suggest that a score that is based on the number of correct responses is

inappropriate for a test that adapts level of difficulty during the test administration. Indeed, it has been argued that scores from computerized adaptive tests might not be comparable to those on standard tests, because the item selection and test taking experience can differ dramatically [48].

Despite concerns related to hardware and software-based irregularities in measurement [52], most cognitive screening efforts can be well-served by the level of precision offered by computer-based assessments. On the other hand, human factors such as spontaneous verbalizations or test taking behaviors are not captured by most current automated examination paradigms. Computerized cognitive tests to date do not offer the richness of qualitative data available from a full neuropsychological examination, and this is not their intent. Future enhancements might address some of these shortcomings and are likely to take advantage of improvements in automated speech recognition, allowing for less reliance on typing or manual input for response, as well as evaluation of quality of speech itself. In this sense, pauses, tone and intonation, variability of responses, and prior patterns of responses might be integrated into the assessment of cognitive function. These new methods will only be possible if the technology behind them is rigorously tested and validated with meaningful outcomes. These developments will evolve, resting on the strengths of computerized testing: the standardization and accuracy of stimulus presentation, measures of response latencies, and automated comparison in real time with an individual's prior performance as well as his or her peers.

Finally, the experience of the user needs to be considered in test development and refinement. In our experience, some difficult computer tests have brought elder users to tears, an outcome often avoided by an experienced examiner who knows how to add a soft touch when needed. Encouragingly, however, others have found that elderly test takers rate computer-based tests as understandable and easy to use [53] and more acceptable than paper and pencil tests [54]. The CANS-MCI is a good example of a computer test that is well-paced and adds a measure of pleasantness to the experience of the user.

Acknowledgements

Supported in part by grants from the NIH: Oregon Alzheimer Disease Center (P30 AG008017) and Oregon Roybal Center for Translational Research on Aging (P30 AG024978).

Dr Webbe has a relationship with Headminder, Inc of New York City, publisher of the CSI. They allow him use of their products for research and clinical work in exchange for the incorporation of his data for the development of norms for their tests. He has not received remuneration from Headminder, Inc. He has published with David Erlanger, who is president of Headminder, and with Tanya Kaushik who is one of their directors.

References

- [1] Golden CJ, Sivan AB. Ethical challenges in neuropsychological test development. In: Bush SS, ed. *A casebook of ethical challenges in neuropsychology*. New York: The Psychology Press; 2005.
- [2] Browndyke JN, Schatz P. Ethical challenges with the use of information technology and telecommunications in neuropsychology. In: Bush SS, ed. *A casebook of ethical challenges in neuropsychology*. New York: The Psychology Press; 2005.
- [3] Schlegel RE, Gilliland K. Development and quality assurance of computer-based assessment batteries. *Arch Clin Neuropsychol* 2007; 22S:S49–61.
- [4] Chalmers I, Altman DG. *Systematic reviews*. London: BMJ Publishing Group; 1995.
- [5] Kabat MH, Kane RL, Jefferson AL, DiPino RK. Construct validity of selected Automated Neuropsychological Assessment Metrics (ANAM) battery measures. *Clin Neuropsychol* 2001;15:498–507.
- [6] Levinson D, Reeves D, Watson J, Harrison M. Automated neuropsychological assessment metrics (ANAM) measures of cognitive effects of Alzheimer's disease. *Arch Clin Neuropsychol* 2005;20:403–8.
- [7] Reeves DL, Winter KP, Bleiberg J, Kane RL. ANAM Genogram: Historical perspectives, description, and current endeavors. *Arch Clin Neuropsychol* 2007;22(Suppl 1):S15–37.
- [8] Kane RL, Roebuck-Spencer T, Short P, Kabat M, Wilken JA. Identifying and monitoring cognitive deficits in clinical populations using automated neuropsychological assessment metrics (ANAM) tests. *Arch Clin Neuropsychol* 2007;22(Suppl 1):S115–26.
- [9] Roebuck-Spencer T, Sun W, Cernich AN, Farmer K, Bleiberg J. Assessing change with the Automated Neuropsychological Assessment Metrics (ANAM): issues and challenges. *Arch Clin Neuropsychol* 2007;22(Suppl 1):S79–87.
- [10] Short P, Cernich A, Wilken JA, Kane RL. Initial construct validation of frequently employed ANAM measures through structural equation modeling. *Arch Clin Neuropsychol* 2007;22(Suppl 1):S15–37.
- [11] Tornatore JB, Hill E, Laboff J, McGann ME. Self-administered screening for mild cognitive impairment: Initial validation of a computerized test battery. *J Neuropsychiatr Clin Neurosci* 2005;17:98–105.
- [12] Robbins TW, James M, Owen AM, Sahakian BJ, McInnes L, Rabbitt PM. Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. *Dementia* 1994;5:266–81.
- [13] Sahakian BJ, Owen AM. Computerized assessment in neuropsychiatry using CANTAB: discussion paper. *J Royal Soc Med* 1992;85: 399–402.
- [14] Fray PJ, Robbins TW, Sahakian BJ. Neuropsychiatric applications of CANTAB. *Int J Geriatr Psychiatr* 1996;11:329–36.
- [15] Lowe C, Rabbitt PM. Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: theoretical and practical issues. *Neuropsychologia* 1998;36:915–23.
- [16] De Luca CR, Wood SJ, Anderson V, Buchanan J, Proffitt TM, Mahony K, et al. Normative data from the Cantab: I—development of executive function over the lifespan. *J Clin Exp Neuropsychol* 2003;25:242–54.
- [17] Fowler KS, Saling MM, Conway EL, Semple JM, Louis WJ. Computerized neuropsychological tests in the early detection of dementia: prospective findings. *J Int Neuropsychol Soc* 1997;3:139–46.
- [18] Fowler KS, Saling MM, Conway EL, Semple JM, Louis WJ. Paired associate learning in the early detection of DAT. *J Int Neuropsychol Soc* 2002;8:58–71.
- [19] De Jager CA, Milwain E, Budge M. Early detection of isolated memory deficits in the elderly: the need for more sensitive neuropsychological tests. *Psychol Med* 2002;32:483–91.

- [20] Gualtieri CT, Johnson LG. Neurocognitive testing supports a broader concept of mild cognitive impairment. *Am J Alzheimers Dis Other Dement* 2005;20:359–66.
- [21] Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery. *CNS Vital Signs. Arch Clin Neuropsychol* 2006;21:623–43.
- [22] Veroff A, Cutler N, Sramek JJ, Prior PL, Mickelson W, Hartman JK. A new assessment for neuropsychopharmacologic research: the Computerized Neuropsychological Test Battery. *J Geriatr Psychiatry Neurol* 1991;4:211–7.
- [23] Cutler NR, Shrotriya RC, Sramek JJ, Veroff AE, Seifert RD, Reich LA, et al. The use of the Computerized Neuropsychological Test Battery (CNTB) in an efficacy and safety trial of BMY 21.502 in Alzheimer's disease. *Ann N Y Acad Sci* 1993;695:332–6.
- [24] Veroff AE, Bodick NC, Offen WW, Sramek JJ, Cutler NR. Efficacy of xanomeline in Alzheimer disease: cognitive improvement measured using the Computerized Neuropsychological Test Battery (CNTB). *Alz Dis Assoc Disord* 1998;4:304–12.
- [25] Simpson PM, Surmon DJ, Wesnes KA, Wilcock GK. The Cognitive Drug Research Computerized Assessment System for demented patients: a validation study. *Int J Geriatr Soc* 1991;6:95–102.
- [26] Mohr E, Walker D, Randolph C, Sampson M, Mendis T. Utility of clinical trial batteries in the measurement of Alzheimer's and Huntington's dementia. *Int Psychogeriatr* 1996;8:397–411.
- [27] Ballard C, O'Brien J, Gray A, Cormack F, Ayre G, Rowan E, et al. Attention and fluctuating attention in patients with dementia with Lewy bodies and Alzheimer disease. *Arch Neurol* 2001;58:977–82.
- [28] DeLepeliere J, Heyrman J, Baro F, Buntinx F. A combination of tests for the diagnosis of dementia had a significant diagnostic value. *J Clin Epidemiol* 2005;58:217–25.
- [29] Collie A, Maruff P, Darby DG, McStephen M. The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *JINS* 2003;9:419–28.
- [30] Falsetti MG, Maruff P, Collie A, Darby D. Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *J Clin Exp Neuropsychol* 2006;28:1095–112.
- [31] Darby D, Maruff P, Collie A, McStephen M. Mild cognitive impairment can be detected by multiple assessments in a single day. *Neurology* 2002;59:1042–6.
- [32] Maruff P, Collie A, Darby D, Weaver-Cargin J, McStephen M. Subtle cognitive decline in mild cognitive impairment (technical document). Australia: CogState Ltd; 2002.
- [33] Cargin JW, Maruff P, Collie A, Masters C. Mild memory impairment in healthy older adults is distinct from normal aging. *Brain Cogn* 2006;60:146–55.
- [34] Erlanger DM, Kaushik T, Broshek D, Freemant J, Feldman D, Festa J. Developments and validations of a web-based screening tool for monitoring cognitive status. *J Head Trauma Rehabil* 2002;17:458–76.
- [35] Lichtenberg PA, Johnson AS, Erlanger DM, Kaushik T, Maddens ME, Imam K, et al. Enhancing cognitive screening in geriatric care: use of an internet-based system. *Int J Healthcare Inf Sys Informatics* 2006;1:47–57.
- [36] Trenkle DL, Shankle WR, Azen SP. Detecting cognitive impairment in primary care: performance assessment of three screening instruments. *J Alzheimers Dis* 2007;11:323–35.
- [37] Shankle WR, Romney AK, Hara J, Fortier D, Dick MB, Chen JM, et al. Methods to improve the detection of mild cognitive impairment. *Proc Natl Acad Sci USA* 2005;102:4919–24.
- [38] Green RC, Green J, Harrison JM, Kutner MH. Screening for cognitive impairment in older individuals. *Arch Neurol* 1994;51:779–86.
- [39] Elwood RW. MicroCog: assessment of cognitive functioning. *Neuropsychol Rev* 2001;11:89–100.
- [40] Johnson JA, Rust JO. Correlational analysis of MicroCog: assessment of cognitive functioning with the Wechsler Adult Intelligence Scale-III for a clinical sample of veterans. *Psychol Rep* 2003;93:1261–6.
- [41] Helmes E, Miller M. A comparison of MicroCog and the Wechsler Memory Scale (3rd ed.) in older adults. *Appl Neuropsychol* 2006;13:28–33.
- [42] Doniger GM, Zucker DM, Schweiger A, Dwolatzky T, Chertkow H, Crystal H, et al. Towards a practical cognitive assessment for detection of early dementia: a 30-minute computerized battery discriminates as well as longer testing. *Curr Alzheimer Res* 2005;2:117–24.
- [43] Dwolatzky T, Whitehead V, Doniger GM, Simon ES, Schweiger A, Jaffe D, et al. Validity of a novel computerized cognitive battery for mild cognitive impairment. *BMC Geriatrics* 2003;3:1–12.
- [44] Dwolatzky T, Whitehead V, Doniger GM, Simon ES, Schweiger A, Jaffe D, et al. Validity of the Mindstreams computerized cognitive battery for mild cognitive impairment. *J Mol Neurosci* 2004;24:33–44.
- [45] Doniger GM, Dwolatzky T, Zucker DM, Chertkow H, Crystal H, Schweiger A, et al. Computerized cognitive testing battery identifies mild cognitive impairment and mild dementia even in the presence of depressive symptoms. *Am J Alzheimers Dis Other Dement* 2006;21:28–36.
- [46] Buchanan T. Online assessment: desirable or dangerous? *Professional Psychology Research and Practice* 2002;33:148–54.
- [47] Butcher JN, Perry JN, Atlis MM. Validity and utility of computer-based test interpretation. *Psychol Assess* 2000;12:6–18.
- [48] Mead AD, Drasgow F. Equivalence of computerized and paper-and-pencil cognitive ability tests: a meta-analysis. *Psychol Bull* 1993;114:449–58.
- [49] Raymond PD, Hinton-Bayre AD, Radel M, Ray MJ, Marsh NA. Test-retest norms and reliable change indices for the MicroCog battery in a healthy community population over 50 years of age. *Clin Neuropsychol* 2006;20:261–70.
- [50] Kane R, Kay GG. Computerized assessment in neuropsychology: a review of tests and test batteries. *Neuropsychol Rev* 1992;3:1–118.
- [51] Schatz P, Browndyke J. Applications of computer-based neuropsychological assessment. *J Head Trauma Rehabil* 2002;17:395–410.
- [52] Cernich AN, Brennana DM, Barker LM, Bleiberg J. Sources of error in computerized neuropsychological assessment. *Arch Clin Neuropsychol* 2007;22:S39–48.
- [53] Fillitt HM, Simon ES, Doniger GM, Cummings JL. Practicality of a computerized system for cognitive assessment in the elderly. *Alzheimer's & Dementia* 2008;4:14–21.
- [54] Collerton J, Collerton D, Yasumichi A, Barrass K, Eccles M, Jagger C, et al. A comparison of computerized and pencil-and-paper tasks in assessing cognitive function in community-dwelling older people in the Newcastle 85+ pilot study. *J Am Geriatr Soc* 2007;55:1630–5.

Appendix

Template for rating test batteries

Subtests (comprehensiveness of domains covered/depth of coverage within domains)

- 1 = Narrow focus, lack of depth
- 2 = Comprehensive but not in depth or narrow focus but adequate depth
- 3 = Comprehensive and in depth

Normative data

- 1 = No data
- 2 = Small sample size of elderly, otherwise okay
- 3 = Adequate sample of elderly

Reliability (inter-rater, test-retest, internal consistency)

- 1 = No data
- 2 = Adequate data, 1 type of reliability
- 3 = >1 type of reliability reported

Validity (content, construct, criterion)

- 1 = No data
- 2 = Adequate data, 1 type of validity
- 3 = >1 type of validity reported

Factor analysis

- 1 = No data
- 3 = Any factor analysis reported

Administration/interface

- 1 = Poorly designed or described interface
- 2 = Reliance on administrator, but good interface
- 3 = Independent administration, self-explanatory with good interface