

Cognitive & Behavioral Assessment

Computer simulations for assessing cognitively intensive instrumental activities of daily living in older adults

Stephen R. Rapp^{a,*}, Ryan T. Barnard^b, Kaycee M. Sink^{c,d}, Dana G. Chamberlain^c,
Valerie Wilson^c, Lingyi Lu^b, Edward H. Ip^b

^aDepartment of Psychiatry and Behavioral Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

^bDepartment of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA

^cDepartment of Internal Medicine (Geriatrics and Gerontology), Wake Forest School of Medicine, Winston-Salem, NC, USA

^dGenentech, San Francisco, CA, USA

Abstract

Introduction: Aim is to evaluate validity, reliability, diagnostic precision, and user acceptability of computer simulations of cognitively demanding tasks when administered to older adults with and without cognitive impairment.

Methods: Five simulation modules were administered to 161 individuals aged ≥ 60 years with no cognitive impairment ($N = 81$), mild cognitive impairment ($N = 52$), or dementia ($N = 28$). Groups were compared on total accuracy and time to complete the tasks (seconds). Receiver operating characteristics were evaluated. Reliability was assessed over one month. Participants rated face validity and acceptability.

Results: Total accuracy ($P < .0001$) and time ($P = .0015$) differed between groups. Test-retest correlations were excellent (0.79 and 0.88, respectively). Area under the curve ranged from good (0.77) to excellent (0.97). User ratings supported their face validity and acceptability.

Discussion: Brief computer simulations can be useful in assessing cognitive functional abilities of older adults and distinguishing varying degrees of impairment.

© 2018 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Cognitive impairment; Assessment; Computer; Simulation; Validation

The prevalence of cognitive impairment is increasing markedly as the number of older Americans increases. Diagnoses of mild cognitive impairment (MCI) and dementia (DM) require the assessment of functional performance in ecologically relevant activities such as managing money, taking medications, navigating social interactions, and self-care [1].

Assessing functional status and decline can be challenging. Self-reports can be unreliable, especially if cognitive abilities are compromised plus individuals may for various reasons choose to conceal or underreport deficits. Reports by proxies are also subject to biases and thus can

be inaccurate [2]. Probably the best method, direct observation of an individual performing relevant activities in their natural environment, is prohibitively expensive. Clinic-based behavioral simulations are useful but often require referrals, therefore adding delays, expense, and inconvenience for patients and families. An optimal assessment strategy would be one that quickly provides relevant information to providers, patients, and families about functional impairments and proficiencies; that is brief and easy-to-administer in a busy clinic; and that has demonstrated reliability, validity, and diagnostic accuracy.

The present study describes an innovative approach to measuring performance in cognitively intensive everyday activities. The SIMulation-Based Assessment of Cognition (SIMBAC) consists of computer tablet-based simulations of

The authors have declared that no conflict of interest exists.

*Corresponding author. Tel.: (336)-716-6995; Fax: (336)-716-6830.

E-mail address: srapp@wakehealth.edu

<https://doi.org/10.1016/j.dadm.2018.01.008>

2352-8729/© 2018 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

five common, cognitively demanding activities—recognizing faces, remembering names, filling a pillbox, using an automated teller machine (ATM), and renewing a medication prescription over the phone. The aim of this study was to assess the validity, reliability, diagnostic precision, ecological relevance, and user acceptability of the SIMBAC modules when administered to older adults with and without independently adjudicated cognitive impairment (MCI, DM).

1. Methods

1.1. Development of SIMBAC modules

Two board-certified geriatricians (K.S. and V.W.) and a clinical geropsychologist (S.R.) identified five instrumental abilities from their clinical experience that were cognitively demanding, relevant to everyday life for most older Americans and could be simulated on a computer. They were as follows: recognizing faces, remembering names, filling a pillbox, using an ATM, and refilling a prescription over the phone. A computer programmer then (R.B.) developed each module to be administered on a computer tablet. The SIMBAC modules are as follows:

Orientation to the computer: An initial orientation module provides written instructions on the use of the com-

puter tablet and exercises using touch and “dragging” features of the tablet. A trained examiner assists the respondent, as needed, and repeats the training trials until the participant can complete them without error.

FACES: Respondents view a digital photo of a human face (varied by gender, age, and race/ethnicity) for 5 seconds. Next, a series of gender-, age-, and race-matched novel facial images [3] are presented together with the target image, and respondents are asked to touch the target image. A practice trial is followed by three successive trials in which the target image is presented with 1, 2, or 3 distractor images.

FACES AND NAMES: Respondents view a series of digital facial images each paired with a unique name for 5 seconds. The next screen presents all the viewed faces but only one name (Fig. 1A). The individual must touch the face that was paired with that name. Six trials are presented with 2, 3, or 4 name-face pairs. As with the FACES module, facial images vary by gender, age, and race/ethnicity (whites, African-Americans, and Hispanic).

PILLBOX: A 7-day pillbox with a.m. and p.m. compartments is presented next to images of three pill containers, each with unique instructions printed on the label (e.g., “Take one tablet with a glass of water daily in the evening.”). Images of three uniquely shaped or colored tablets/capsules are shown next to each bottle. Participants are

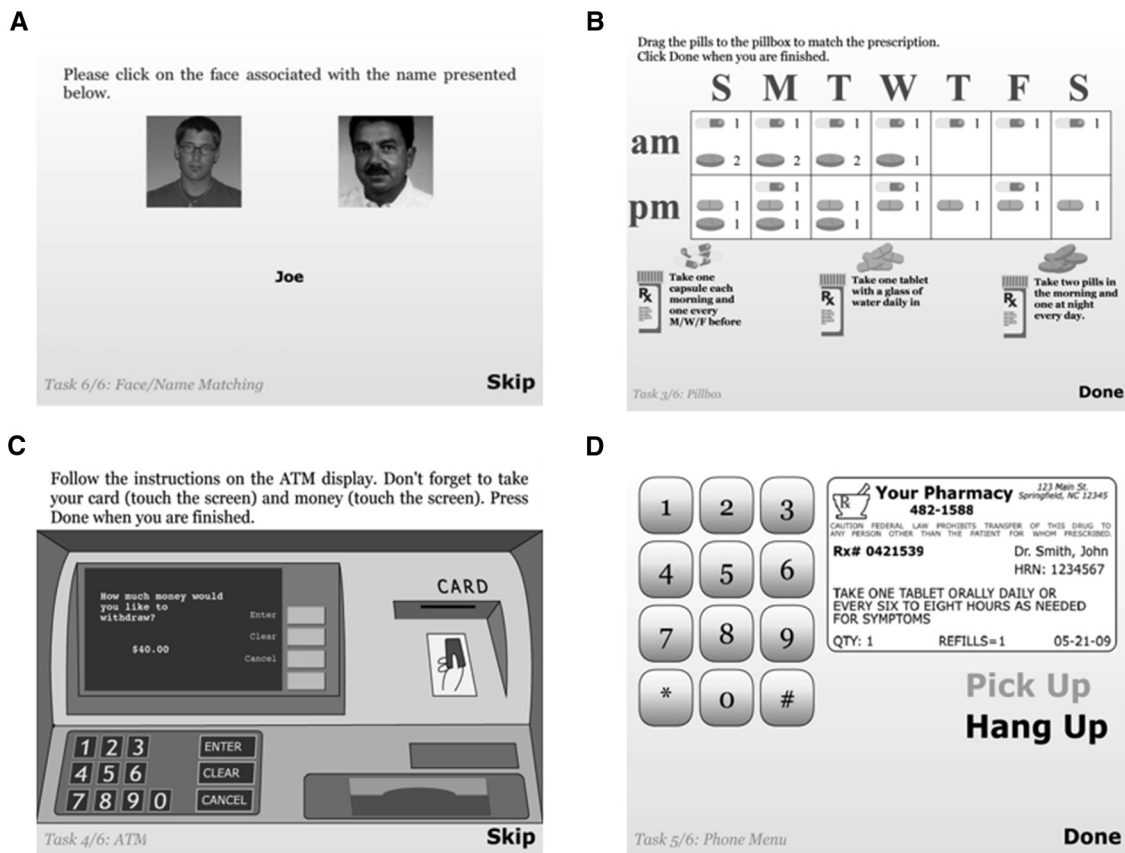


Fig. 1. Screenshots of SIMBAC modules. Module: (A) FACES & NAMES, (B) PILLBOX, (C) ATM, (D) PHONE. Abbreviation: SIMBAC, SIMulation-Based Assessment of Cognition.

instructed to fill the pillbox by touching and dragging the pills to the correct compartments (Fig. 1B).

Automated teller machine: The initial screen informs respondents they will be asked to withdraw money from an ATM by inserting an ATM card, typing in a personal identification number, specifying a specific amount of cash and removing their money and ATM card. Next, they are instructed how to insert and withdraw the ATM card and to take the money by touching and dragging the icons. The test screen shows an ATM screen and keypad along with a four-digit personal identification number and written instructions to withdraw \$40 from their checking account, not to request a receipt and to remember to retrieve their card and money. The screen also shows an ATM card (Fig. 1C). As with actual ATMs, respondents can return to the instruction screen by pressing a "return" button.

Automated prescription renewal (PHONE): Respondents are instructed to renew a medication prescription over phone at the pharmacy using the information printed on a simulated pill bottle label and the automated telephone messaging system. The test screen presents a telephone keypad, a label with all necessary information printed on it and an image of a phone receiver (Fig. 1D). Participants touch the phone icon and then hear a ringing tone followed by an automated voice recorder with step-by-step instructions.

Administration: Respondents were seated at a desk in a quiet room next to the trained technician. The modules were presented in the same order on a Motorola XOOM 10.1 inch computer tablet. Total testing time was approximately 10 minutes.

Scoring: Response parameters judged as most useful for clinicians, individuals, and family members were accuracy and speed. A SIMBAC total accuracy (TA) score is the sum of the five module accuracy scores. A module accuracy score is calculated by awarding two points if the module is completed without any errors, 1 point if it is completed with ≥ 1 error or 0 points if the individual could not complete it. TA scores can range from 0 to 10. Time (seconds) to complete each module was summed to yield the SIMBAC total time (TT) score.

1.2. Validity and reliability assessment

Criterion validity was evaluated by comparing SIMBAC accuracy and time scores across three groups of older adults who were independently determined to have no cognitive impairment (NI), MCI, or DM. It was predicted that the TA would be highest in the NI group and lowest in the DM group and the TT would be shortest in the NI group and longest in the DM group. TA and TT scores were also correlated with total scores from a standardized proxy-completed Lawton Instrumental Activity of Daily Living (IADL) rating scale. Ecological validity was assessed by participants' responses to the question, "How often have you (e.g., used an ATM) in the past year?" (not even once = 1, several times = 2, monthly = 3, weekly

or more often = 4). Face validity was assessed by responses to the question, "How realistic was this computer task?" (very unrealistic = 1, somewhat unrealistic = 2, somewhat realistic = 3, very realistic = 4). User acceptability was assessed by responses to the question, "Would you complete (the module) if your doctor requested it?" (no/yes). Test-retest reliability was assessed by examining the intraclass correlations of total accuracy score and total time score across two separate administrations approximately one month apart in a subset ($n = 66$) of participants selected from each group.

1.2.1. Participants

One hundred sixty-two English-speaking, noninstitutionalized adults who were aged ≥ 60 years were recruited from (1) the Wake Forest University Baptist Medical Center Memory Assessment Clinic and (2) the surrounding metropolitan area with local newspaper and newsletter advertisements. Exclusion criteria included significantly impaired vision, hearing, or motor performance or history of stroke, Parkinson's disease, schizophrenia, bipolar disorder, or substance dependence and persons with reversible cognitive impairment. Current diagnosis of depression was permitted if the individual was on stable doses of medication (past 60 days) and not experiencing an acute episode. Use of cognitive enhancing medications was permitted if the regimen was stable for at least 30 days. Individuals with severe cognitive impairment (Mini-Mental State Examination [MMSE] ≤ 11) were excluded.

1.2.2. Procedure

After providing informed consent, all participants were evaluated following a specific clinical protocol used in the Memory Assessment Clinic that included standardized neurocognitive testing (MMSE [4], Rey Auditory Verbal Learning Test [5], Trail Making Test [6], Verbal Fluency–Animals [7], Digit Symbol Coding [8], 30-item Boston Naming Test [9], Digit Span [8], Logical Memory [10], and 15-item Geriatric Depression Scale [11]); a history and physical examination by a geriatric specialist; and an interview with a family member, which included administration of standardized questionnaires (Neuropsychiatric Interview Questionnaire [12], Katz Activity of Daily Living Scale [13], and IADL [14] Scale). Generally, cognitive tests scores ≥ 1.5 standard deviation units worse than in age- and education-appropriate normative groups were considered abnormal. Optional laboratory tests and brain imaging were administered as needed. The board-certified geriatricians (K.M.S. and V.W.) then assigned the diagnosis of NI, MCI, or DM according to established criteria [15,16] using all available data except the SIMBAC results. The SIMBAC modules were administered to each participant at the end of the neurocognitive testing session.

1.3. Analysis

The validity of the instrument was assessed by comparing participants' mean performance (accuracy and time) for

each module and for the summary TA and TT scores across the three clinical groups using chi-square and univariate analysis of variance tests, respectively. Results unadjusted and adjusted for age and education are reported. Multivariate analysis of variance with both TA and TT scores is reported as well. Tukey and Bonferroni-adjusted post hoc pairwise comparisons were performed for individual modules and summary scores. Spearman correlations between IADL and both TA and TT scores were used to assess convergent validity. Ecological and face validity and user acceptability were reported in terms of percentage of participants who affirmed that they engaged in the activity in the past year, considered the module realistic, and attested that they would do the simulations for if their doctor asked. Test-retest reliability of the TA and TT scores was assessed by intraclass correlation coefficient. The diagnostic classification accuracy of the SIMBAC TA score was assessed with receiver operating characteristics analysis. To understand how well SIMBAC classified participants with differing levels of cognitive impairment, two-group comparisons were examined—NI versus MCI + DM, NI versus MCI, NI versus DM, NI + MCI versus DM, and MCI versus DM. For each contrast, area under the curve (AUC) classification accuracy, sensitivity, specificity, and optimal cutoff value that maximized Youden's index (sensitivity + specificity - 1) were calculated. For comparison, the classification accuracy of the MMSE (<24 = impaired, >24 = unimpaired) and a mean cognitive composite score created from standardized scores on the RAVLT-Delayed Recall, Boston Naming Test, Digit Span, Digit Symbol Coding, and Trail Making Test-Part B were calculated.

Table 1
Demographic and clinical characteristics of the no impairment, mild cognitive impairment, and dementia groups

Variable	Total sample, N = 161	No impairment, N = 81	Mild cognitive impairment, N = 52	Dementia, N = 28	P value
Age (years), mean (SD), [range, years]	73.4 (7.9), [60–95]	70.5 (7.8), [60–94]	75.6 (6.8), [61–95]	77.5 (7.1), [62–87]	<.001
Gender (female), frequency (%)	99 (61)	52 (64)	31 (60)	16 (57)	.76
Education > HS, frequency (%)	132 (81)	77 (95)	41 (79)	14 (50)	<.001
Race/ethnicity (white), frequency (%)	131 (90)	64 (86)	42 (93)	25 (93)	.48
Mini-Mental State Examination, mean (SD), [range]	27.1 (3.5), [13–30]	29.2 (1.2), [24–30]	26.4 (2.7), [18–30]	22.5 (4.1), [13–28]	<.001
Family history of dementia (yes), frequency (%)	79/131 (56)	41/71 (58)	22/44 (50)	16/26 (62)	.6
GDS, mean (SD), [range]	2.2 (2.4), [0–12]	1.5 (1.7), [0–7]	2.6 (2.6), [0–11]	3.7 (3.0), [0–12]	<.001
IADL Scale, mean (SD), [range]	18.3 (3.8), [8–21]	20.6 (1.2), [15–21]	17.1 (3.9), [10–21]	13.0 (3.1), [8–19]	<.001
ADL Scale, mean (SD), [range]	17.7 (1.0), [12–18]	17.9 (0.4), [16–18]	17.9 (0.6), [15–18]	17.0 (1.9), [12–18]	<.001
RAVLT Immediate Recall, mean (SD)	34.1 (12.1)	41.6 (9.3)	29.5 (9.3)	19.7 (5.9)	<.001
RAVLT-Delayed Recall, mean (SD)	5.0 (4.0)	7.3 (3.3)	3.3 (2.9)	0.6 (1.4)	<.001
TMT, Part A (seconds), mean (SD)	43.4 (33.3)	30.8 (10.7)	47.5 (27.2)	73.6 (60.0)	<.001
TMT, Part B (seconds), mean (SD)	101.3 (71.0)	82.9 (38.0)	119.7 (79.7)	121.0 (110.0)	.0042
Word Fluency, mean (SD)	18.2 (6.5)	22.3 (4.5)	15.9 (5.6)	10.5 (3.7)	<.001
Digit Symbol Coding, mean (SD)	49.2 (19.5)	61.6 (13.8)	40.6 (12.5)	28.0 (19.2)	<.001
Boston Naming Test, mean (SD)	25.2 (5.0)	27.9 (2.0)	23.8 (4.3)	19.5 (6.8)	<.001
Digit Span, mean (SD)	10.8 (2.0)	11.5 (1.9)	10.4 (1.8)	9.5 (2.0)	<.001
Logical Memory 1, mean (SD)	33.4 (14.2)	43.0 (9.5)	27.5 (10.8)	15.0 (7.6)	<.001
Logical Memory 2, mean (SD)	18.4 (11.1)	26.4 (6.9)	13.0 (7.8)	4.0 (5.7)	<.001

Abbreviations: GDS, Geriatric Depression Scale-Short Form; IADL, Lawton Independent Activities of Daily Living; ADL, Katz Activities of Daily Living; RAVLT, Rey Auditory Verbal Learning Test; TMT, Trail Making Test.

2. Results

One hundred sixty-one participants had evaluable SIMBAC data. Most study participants self-identified as non-Hispanic white (90%), were more likely to be female (61%), with greater than a high school education (81%), and a mean age of 73.4 (7.9) years. Approximately half (56%) reported a family history of DM (Table 1).

The three clinical groups differed significantly in age ($P < .001$) and education ($P < .001$), but not sex, race/ethnicity, or family history of DM ($P > .05$). As expected, there were significant group differences in cognitive test performance, neuropsychiatric symptoms, depressive symptom severity, and Katz Activity of Daily Living and IADL performance (Table 1). Within the DM group, 20 (71%) participants were subclassified with Alzheimer's type, 4 (14%) with Lewy Body type, and 2 (7%) with frontotemporal type and 2 (7%) were missing data.

All study participants achieved errorless performance on the tablet orientation module before attempting the other modules indicating adequate proficiency.

2.1. Criterion validity

Table 2 shows the SIMBAC module and TA and time performance by clinical group. For each module, the proportion of respondents who performed errorlessly significantly differed across groups ($P_s < .005$) and differences remained significant after adjusting for age and education (all $P_s < .001$). As predicted, the NI group performed the best and the DM group performed the poorest on all modules. Means for time (seconds) to complete each module also were significantly different for all five modules ($P < .05$)

with the NI group having the shortest times and the DM group having the longest. Differences remained significant after adjustment for age and education for FACES, ATM, and PILLBOX ($P < .05$) but not for PHONE ($P = .78$) or FACES & NAMES ($P = .07$). The adjusted multivariate analysis of variance for the TA and TT was significant (Pillai's Trace $F [4,316] = 29.25, P < .001$) as were the adjusted univariate analyses of variance for TA ($F [2,156] = 43.3, P < .001$) and TT ($F [2,156] = 6.8, P < .0015$), with the best performance in the NI group and poorest in the DM group. Bonferroni and Tukey pairwise comparisons for individual module mean accuracy scores revealed that all modules significantly distinguished NI from MCI participants ($P < .05$) and the PHONE, ATM, and PILLBOX module scores distinguished MCI from DM participants ($P < .05$, Table 2). The mean time scores for FACES & NAMES, PILLBOX, and ATM also significantly distinguished NI from MCI participants ($P < .05$) and FACES, ATM, and PHONE significantly distinguished the MCI from DM ($P < .05$). Finally, mean overall TA and TT scores and Tukey pairwise comparisons were significantly different across groups ($P < .05$).

Participants who reported not currently engaging in the module's activity in real life were less accurate on the ATM ($P < .0001$) and PILLBOX ($P < .02$) tasks compared to those who reported engaging in these tasks. No difference between engagers and nonengagers was observed on the PHONE task ($P > .05$). Engagers performed significantly faster on the

ATM ($P < .004$) task, but not on the other tasks ($P > .05$) (Supplementary Table 1). The FACES and FACES & NAMES tasks were not included in this analysis because it was presumed that these activities remain ongoing throughout life.

The correlations between the TA and TT and the informant-provided IADL score were $\rho = 0.68$ ($P < .0001$) and $\rho = -0.45$ ($P < .0001$), respectively, with higher accuracy and faster SIMBAC performance associated with better proxy-reported IADL functioning.

2.2. Ecological and face validity and user acceptability

The percentages of participants who had engaged in the activities featured in the SIMBAC modules during the past year were as follows: remembering faces and names, 93%; filling a pillbox, 59%; using an ATM, 57%; and renewing a prescription over phone, 75% (overall mean = 71%). Not surprisingly, a higher proportion of DM participants reported not currently engaging in filling a pillbox, using an ATM and renewing a prescription. Respondents uniformly rated the modules as realistic (range: 94%–99%) and almost all (93% for each module) indicated their willingness to complete SIMBAC tasks if their doctor requested it.

2.3. Reliability

Sixty-six participants drawn from all three groups completed the SIMBAC modules a second time on average of 3.51 (1.3)

Table 2

Group comparison of performance on SIMBAC modules and summary scores for no impairment, mild cognitive impairment, and dementia groups

SIMBAC module/parameter	Clinical groups			Unadjusted test P value	Adjusted test*
	No impairment, N = 81	Mild cognitive impairment, N = 52	Dementia, N = 28		
Practice					
Successful completion, frequency (%)	81 (100)	52 (100)	28 (100)	NA	NA
FACES					
Errorless performance, frequency (%)	68 (84)	29 (56) [†]	10 (36) [‡]	<.0001	$\chi^2 = 14.1, P = .0008$
Time to complete (seconds), mean (SD)	113.4 (15.7)	115.9 (19.6)	141.4 (42.2) ^{‡,§}	<.0001	$F(2,156) = 10.4, P < .0001$
FACES & NAMES					
Errorless performance, frequency (%)	18 (22)	4 (8)	0 (0) [‡]	.0029	$\chi^2 = 6.9, P = .03$
Time to complete (seconds), mean (SD)	145.4 (10.3)	151.9 (16.1) [†]	158.0 (15.1) [‡]	<.0001	$F(2,156) = 2.7, P = .07$
PILLBOX					
Errorless performance, frequency (%)	57 (70)	15 (29) [†]	2 (7) [‡]	<.0001	$\chi^2 = 23.9, P < .0001$
Time to complete (seconds), mean (SD)	125.8 (40.5)	170.1 (81.6) [†]	192.9 (106.6) [‡]	<.0001	$F(2,156) = 3.2, P = .04$
ATM					
Errorless performance, frequency (%)	65 (80)	29 (56) [†]	2 (7) ^{‡,§}	<.0001	$\chi^2 = 29.9, P < .0001$
Time to complete (seconds), mean (SD)	87.4 (25.5)	111.8 (40.1) [†]	137.4 (58.3) ^{‡,§}	<.0001	$F(2,156) = 8.9, P = .0002$
PHONE					
Errorless performance, frequency (%)	76 (94)	40 (77) [†]	9 (32) ^{‡,§}	<.0001	$\chi^2 = 25.2, P < .0001$
Time to complete (seconds), mean (SD)	106.1 (28.0)	121.0 (40.9)	128.6 (83.5)	.045	$F(2,156) = 0.3, P = .78$
Total accuracy, mean (SD)	8.5 (1.1)	6.9 (1.7) [†]	5.0 (1.5) ^{‡,§}	<.0001	$F(2,156) = 43.3, P < .0001$
Total time, mean (SD)	578.0 (85.6)	670.7 (130.3) [†]	758.3 (224.3) ^{‡,§}	<.0001	$F(2,156) = 6.78, P = .00015$

Abbreviations: ATM, automated teller machine; SIMBAC, SIMulation-Based Assessment of Cognition; MCI, mild cognitive impairment.

NOTE. Pairwise comparison used Tukey method (continuous) and Bonferroni adjustment (categorical). TA range = 0–10; TT range: 0 to infinity.

*Adjusted for age and education; for pairwise comparisons.

[†]Indicates $P < .05$ significance between MCI and no impairment groups.

[‡]Indicates $P < .05$ significance between dementia and no impairment groups.

[§]Indicates $P < .05$ significance between dementia and MCI groups.

Table 3
Results of receiver operating characteristics analyses for SIMBAC total accuracy score

Parameter	NI versus MCI	NI versus DM	NI + MCI versus DM	NI versus MCI/DM	MCI versus DM
Optimal cutoff	8.5	6.5	6.5	7.5	5.5
AUC	0.77	0.97	0.91	0.84	0.82
Accuracy	67%	94%	84%	76%	75%
Sensitivity	79%	86%	86%	70%	71
Specificity	58%	75%	84%	82%	77
MMSE accuracy (cut point <24)	65%	83%	50%	63%	69% (cut point <20)
Composite score (cut point < -0.33)	71%	91%	83%	72%	62% (cut point < -0.65)

Abbreviations: NI, no impairment; MCI, mild cognitive impairment; DM, dementia; AUC, area under the curve; MMSE, Mini-Mental State Examination; SIMBAC, SIMulation-Based Assessment of Cognition.

NOTE. Composite score = average of five standardized score from RAVLT-Delayed Recall, Boston Naming Test (BNT), Digit Span Test (total of backward and forward; DST), Digit Symbol Coding (DSC), and Trail Making Test-Part B (TMT-B).

weeks after their first administration. The intraclass correlation for test-retest was 0.79 for the TA and 0.88 for the TT scores indicating good-to-excellent temporal stability.

2.4. Diagnostic accuracy

Table 3 details the results for the receiver operating characteristics analyses using the SIMBAC TA score to predict group membership. AUC was greatest when distinguishing NI participants from DM participants (AUC: 0.97), with almost perfect classification (97%). The AUCs for other group comparisons ranged from 0.77 for NI versus MCI (accuracy = 67%) to 0.91 for NI + MCI versus DM (accuracy = 84%). For each group comparison, the classification rate was higher for SIMBAC than for the MMSE and for the cognitive composite score (see Table 3).

3. Discussion

A diagnosis of MCI and DM requires an assessment of functional abilities [15,16]. Reliance on patients' self-report of their abilities may be unreliable and proxy reports can be inaccurate. Questionnaires rely on relatively coarse qualitative judgments of complex tasks by friends or family members and may lack sensitivity to subtler losses of function. Direct systematic assessment of a person's functional abilities is generally not feasible.

SIMBAC, a brief (10 minutes) and easy-to-administer set of computerized simulations, was developed to provide useful information in the assessment of the functional capabilities of older adults. Data presented here provide preliminary support for its validity, reliability, diagnostic precision, and acceptability to users. In this sample of predominantly white, well-educated older men and women, performance on the modules significantly distinguished the three clinical groups, was consistent over a 1-month period, correlated with proxy-reported functional status, correctly classified members according to their level of cognitive impairment adequately to very well, and was acceptable to participants.

Computer applications have been used with older adults to train cognitive abilities [17-19], assess cognitive

performance [20-23], and measure driving skills [24,25], but we found no reports of computer simulations assessing a variety of cognitively demanding instrumental activities of daily living designed for use with cognitively impaired older adults. Ruse et al. [26] developed the Virtual Reality Functional Capacity Assessment Tool, an interactive gaming-based measure of functional capacity designed specifically for individuals with schizophrenia. Czaja et al. [27] evaluated the psychometric properties of a battery of three computer simulations to assess functional capacity of adults with schizophrenia. It included an ATM and a telephone prescription renewal tasks, similar to the SIMBAC battery and a form completion task. They demonstrated the feasibility of using simulations for this purpose. Our findings extend the use of computer simulations to assess functional capabilities of cognitively impaired older adults.

Although the SIMBAC accuracy scores were significantly different between clinical groups for each module, the FACES & NAMES module showed floor effects; only 22% of NI participants could complete this task without errors. By contrast, the FACES, PILLBOX, ATM, and PHONE modules all showed a wide range of scores and large between-group differences in both accuracy and time. To explore the relative contribution of the FACES & NAMES task, we repeated our AUC analysis after removing FACES & NAMES from the total accuracy score (Supplementary Table 2). Doing so reduced sensitivity for distinguishing MCI from NI from 79% to 58%, more than any of the other tasks. Thus, FACES & NAMES may contribute importantly to the discrimination between individuals at the higher levels of cognitive performance, which is valuable for early detection of abnormal cognitive decline. Including more finely graded trials might strengthen the module.

Approximately 40% of respondents reported that they had not used an ATM or filled a pillbox in the past year. This may be because they were having difficulties with these tasks and responsibility had been taken over by someone else. We did find that participants in the DM group were more likely not to have performed these tasks. We must determine yet whether individuals from more demographically, socioeconomically, and ethnically diverse populations normally engaged in these tasks.

The TA score demonstrated good-to-excellent classification of participants into their diagnostic groups indicating SIMBAC performance might help when making a diagnosis. Classification was almost perfect (94%) when distinguishing older adults without cognitive impairment from those with mild DM (mean MMSE = 22.5, SD = 4.1). We were encouraged that the TA score outperformed the MMSE, the widely used measure of global cognitive function as well as a cognitive composite score derived from our cognitive battery.

The SIMBAC TA and TT scores were significantly and moderately correlated with proxy-reported IADL questionnaire scores suggesting they measure somewhat different phenomena and further support the convergent validity of SIMBAC. Moreover, proxy reports possess inherent limitations (e.g., proxy biases, inadequate observations of performance) not found in direct assessment of performance.

These computerized modules can yield additional useful information about cognitive functioning, such as response latency, error rates, and number of repeated attempts. We chose accuracy and time for this study because of their easy interpretation by clinicians, patients, and family members. In addition to its potential role in the measurement of functional abilities and diagnosis, SIMBAC could be used in conversations with patients and family members to better describe functional problems or to identify where compensatory strategies might be needed. SIMBAC might also be useful as an outcome measure to assess the efficacy of an intervention or impairment progression. Neuroimaging and biomarker studies correlating performance on the SIMBAC tasks with brain activity and neurophysiology could yield additional useful information.

There are limitations to this study. The five modules do not represent the full range of activities commonly engaged in by older adults. Additional modules measuring other cognitively intensive activities (e.g., driving) could be developed to improve its comprehensiveness, but doing so would come at the expense of SIMBAC's conciseness. In addition, performance on these modules could vary depending on particular neuropathology(ies). The majority of our DM cases were classified as of the Alzheimer's type, and the study lacked statistical power to test SIMBAC performance differences across them. Our sample was small, predominantly white, and well educated, and thus not representative of all older Americans. Whether the tasks portrayed in these modules are relevant to culturally or socioeconomically diverse populations remains to be determined. The reliability of SIMBAC performance was tested over a relatively short 1-month interval. Strengths of the study include the innovativeness of the computer simulations; their brevity; their ease of administration, scoring, and interpretation; their diagnostic accuracy; and their acceptability to older adults of varying cognitive ability.

In conclusion, the present study provides encouraging preliminary evidence that these novel computerized

simulations of cognitively intensive activities of daily living can be an efficient, reliable, and valid assessment tool when used with older adults of varying cognitive abilities.

Acknowledgments

Authors' contributions: Conceptualization of idea was performed by S.R.R., K.M.S., E.H.I., R.T.B., and V.W. Data collection was carried out by S.R.R., K.M.S., V.W., and D.G.C. Data analysis was carried out by S.R.R., E.H.I., L.L., and R.T.B. Writing of article was performed by S.R.R., E.H.I., K.M.S., V.W., R.T.B., and D.G.C. Editing of article was performed by S.R.R., E.H.I., R.T.B., K.M.S., V.W., D.G.C., and L.L.

This work was supported by a grant from the Alzheimer's Association (DNCFI-12-241602).

Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.dadm.2018.01.008>.

RESEARCH IN CONTEXT

1. Systematic review: Assessment of functional abilities is a requirement for a clinical diagnosis of mild cognitive impairment or dementia and useful in developing treatment planning. Current methods of assessing functional abilities have significant limitations. We evaluated the validity, reliability, diagnostic precision, and user acceptability of a brief, easy-to-administer, set of five computer simulations of cognitively intensive functional abilities in a sample of older adults with no cognitive impairment, mild cognitive impairment, and mild dementia.
2. Interpretation: We found that performance (accuracy, time to complete the task) on the computer simulations significantly distinguished the clinical groups, was reliable over a one-month period, correlated with proxy-reported functional abilities, showed good-to-excellent receiver operating characteristics, and demonstrated high user acceptability. This provided preliminary support for the utility of computer simulations for evaluating functional abilities of older adults.
3. Future directions: Additional research will focus on the predictive validity and sensitivity to change of the performance on the simulations; their cost-efficiency when used in the clinic; and the relationship between performance and disease markers.

References

- [1] American Psychiatric Association. Task Force on DSM-IV. Diagnostic and statistical manual of mental disorders: DSM-IV. 4th ed. Washington, DC: American Psychiatric Association; 1994. xxvii, p. 886.
- [2] Loewenstein DA, Arguelles S, Bravo M, Freeman RQ, Arguelles T, Acevedo A, et al. Caregivers' judgments of the functional abilities of the Alzheimer's disease patient: a comparison of proxy reports and objective measures. *J Gerontol B Psychol Sci Soc Sci* 2001; 56:P78–84.
- [3] Minear M, Park DC. A lifespan database of adult facial stimuli. *Behav Res Methods Instrum Comput* 2004;36:630–3.
- [4] Folstein MF, Folstein SE, McHugh PR. 'Mini Mental State': a practical method for grading the cognitive state of patients for the clinician. *J Psychiatry* 1975;12:189–98.
- [5] Strauss E, Sherman EMS, Spreen O, Spreen O. A compendium of neuropsychological tests: administration, norms, and commentary. 3rd ed. Oxford; New York: Oxford University Press; 2006. xvii, p. 1216.
- [6] Reitan R. Trail making test. Manual for administration and scoring. Tucson, AZ: Reitan Neuropsychological Laboratory; 1992.
- [7] Lesak MD. Neuropsychological assessment. 3rd ed. New York: Oxford University Press; 1997:1997.
- [8] Wechsler D. Wechsler Adult Intelligence Scale-III (WAIS-III). New York: Psychological Corporation/Harcourt, Inc; 1996.
- [9] Kaplan E, Goodglass H, Weintraub S. The Boston Naming Test. 2nd ed. Philadelphia, PA: Lea & Febiger; 1983.
- [10] Wechsler D. The Wechsler Memory Scale-3rd Edition (WMS-III). New York: Psychological Corporation, Harcourt, Inc; 1996.
- [11] Yesavage JA. Geriatric depression scale. *Psychopharmacol Bull* 1988; 24:709–11.
- [12] Kaufer DI, Cummings JL, Ketchel P, Smith V, MacMillan A, Shelley T, et al. Validation of the NPI-Q, a brief clinical form of the Neuropsychiatric Inventory. Index of ADL Studies of Illness in the Aged. The Index of Adl: A Standardized Measure of Biological and Psychosocial Function Assessment of older people: self-maintaining and instrumental activities of daily living. *J Neuropsychiatry Clin Neurosci* 2000;12:233–9.
- [13] Katz S, Akpom CA. A measure of primary sociobiological functions. *Int J Health Serv* 1976;6:493–508.
- [14] Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 1969;9:179–86.
- [15] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:270–9.
- [16] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:263–9.
- [17] Bozoki A, Radovanovic M, Winn B, Heeter C, Anthony JC. Effects of a computer-based cognitive exercise program on age-related cognitive decline. *Arch Gerontol Geriatr* 2013;57:1–7.
- [18] Lee GY, Yip CCK, Yu ECS, Man DWK. Evaluation of a computer-assisted errorless learning-based memory training program for patients with early Alzheimer's disease in Hong Kong: a pilot study. *Clin Interv Aging* 2013;8:623–33.
- [19] Rafii M, Taylor C, Coutinho A, Kim K, Galasko D. Comparison of the memory performance index with standard neuropsychological measures of cognition. *Am J Alzheimers Dis* 2011;26:235–9.
- [20] Rabuffetti M, Ferrarin M, Spadone R, Pellegatta D, Gentileschi V, Vallar G, et al. Touch-screen system for assessing visuo-motor exploratory skills in neuropsychological disorders of spatial cognition. *Med Biol Eng Comput* 2002;40:675–86.
- [21] Saxton J, Morrow L, Eschman A, Archer G, Luther J, Zuccolotto A. Computer assessment of mild cognitive impairment. *Postgrad Med* 2009;121:177–85.
- [22] Jacova C, McGrenere J, Lee HS, Wang WW, Le Huray S, Corenblith EF, et al. C-TOC (Cognitive Testing on Computer) investigating the usability and validity of a novel self-administered cognitive assessment tool in aging and early dementia. *Alzheimer Dis Assoc Disord* 2015;29:213–21.
- [23] Wild K, Howieson D, Webbe F, Seelye A, Kaye J. Status of computerized cognitive testing in aging: a systematic review. *Alzheimers Dement* 2008;4:428–37.
- [24] Lew HL, Poole JH, Lee EH, Jaffe DL, Huang HC, Brodd E. Predictive validity of driving-simulator assessments following traumatic brain injury: a preliminary study. *Brain Inj* 2005;19:177–88.
- [25] Nef T, Muri RM, Bieri R, Jager M, Bethencourt N, Tarnanas I, et al. Can a novel web-based computer test predict poor simulated driving performance? A pilot study with healthy and cognitive-impaired participants. *J Med Internet Res* 2013;15:139–51.
- [26] Ruse SA, Davis VG, Atkins AS, Krishnan KRR, Fox KH, Harvey PD, et al. Development of a virtual reality assessment of everyday living skills. *J Vis Exp* 2014;86, <https://doi.org/10.3791/51405>.
- [27] Czaja SJ, Loewenstein DA, Lee CC, Fu SH, Harvey PD. Assessing functional performance using computer-based simulations of everyday activities. *Schizophr Res* 2017;183:130–6.