Assessment of Age-Related Differences in Functional Capacity Using the Virtual Reality Functional Capacity Assessment Tool (VRFCAT)

A.S. Atkins[1], **I. Stroescu**[1], **N.B. Spagnola**[1], **V.G. Davis**[1], **T.D. Patterson**[2], **M. Narasimhan**[3], **P.D. Harvey**[4], and **R.S.E. Keefe**[1,5]

[1]NeuroCog Trials, Inc

[2]University of California San Diego

[3]University of South Carolina

[4]University of Miami Miller School of Medicine

[5]Duke University Medical Center

Abstract

Clinical trials for primary prevention and early intervention in preclinical AD require measures of functional capacity with improved sensitivity to deficits in healthier, non-demented individuals. To this end, the Virtual Reality Functional Capacity Assessment Tool (VRFCAT) was developed as a direct performance-based assessment of functional capacity that is sensitive to changes in function across multiple populations. Using a realistic virtual reality environment, the VRFCAT assesses a subject's ability to complete instrumental activities associated with a shopping trip. The present investigation represents an initial evaluation of the VRFCAT as a potential co-primary measure of functional capacity in healthy aging and preclinical MCI/AD by examining test-retest reliability and associations with cognitive performance in healthy young and older adults. The VRFCAT was compared and contrasted with the UPSA-2-VIM, a traditional performance-based assessment utilizing physical props. Results demonstrated strong age-related differences in performance on each VRFCAT outcome measure, including total completion time, total errors, and total forced progressions. VRFCAT performance showed strong correlations with cognitive performance across both age groups. VRFCAT Total Time demonstrated good test-retest reliability (ICC=.80 in young adults; ICC=.64 in older adults) and insignificant practice effects, indicating the measure is suitable for repeated testing in healthy populations. Taken together, these results provide preliminary support for the VRFCAT as a potential measure of functionally relevant change in primary prevention and preclinical AD/MCI trials.

Keywords

Functional capacity; prodromal AD; aging; functional outcome; meaningful change

Introduction

Reliable evaluation of both cognitive performance and functional capacity are critical to the effective assessment of mental health in aging individuals at increased risk for Mild Cognitive Impairment (MCI) and Alzheimer's disease (AD). Impairments in cognition and function are among the diagnostic criteria for AD (1). Although many aspects of functional capacity are preserved in MCI, including the ability to perform basic activities of daily living (ADLs), MCI is commonly associated with more subtle deficits in instrumental ADLs (IADLs) such as shopping, navigating public transportation and cooking (2-4). As such, reliable assessment of functional abilities and change over time is of considerable importance to tracking decline in both AD and MCI.

The need for improved precision in the assessment of functional capacity is well-recognized in the literature (2, 5). While assessment of cognitive decline is largely standardized with the use of performance-based neuropsychological instruments, standard assessments of functioning such as the Alzheimer's Disease Cooperative Study–Activities of Daily Living (ADCS-ADL) (6), and Functional Activities Questionnaire (FAQ) (7) rely on subjective informant reports. Although such reports can provide valuable insight in many cases, there is evidence for caregiver bias in subjective reporting (8-10) and the psychometric reliability of these scales has been difficult to establish (2, 5). Finally, informant report measures often lack sensitivity to subtle functional deficits in non-demented preclinical and prodromal MCI/AD (11, 12). Given increasing interest in clinical trials for primary prevention and early intervention in preclinical AD, there is a growing need for measures of functional capacity with improved sensitivity to functional deficits in healthier, non-demented individuals (see (2)).

Measures of functional capacity tailored to the preclinical population should focus on assessment of complex IADLs that are most sensitive to early decline. In order to be useful for clinical trials, these measures must demonstrate a variety of psychometric and practical criteria including good test-retest reliability, minimal practice effects, correlations with cognitive measures, and correlations with functional outcomes (13, 14). They should also be practical for clinical investigators and clinicians, and tolerable for participants. When possible, multiple forms should be available to allow for repeated testing.

The Virtual Reality Functional Capacity Assessment Tool (VRFCAT) was developed as a direct performance-based assessment of functional capacity that meets these requirements and is appropriate for use in multiple populations, including preclinical AD and MCI. Using a realistic virtual environment, the VRFCAT assesses a subject's ability to complete instrumental activities associated with a shopping trip, including searching the pantry at home, making a shopping list, taking the correct bus to the grocery store, shopping in the store, paying for groceries, and returning home. In previous studies, the VRFCAT has

demonstrated high test-retest reliability and shown sensitivity to functional impairment (16, 17).

The primary aim of the present study was to evaluate whether the VRFCAT demonstrates the essential psychometric and practicality characteristics required of a functional capacity measure intended for use in a healthy aging population such as that targeted in a primary prevention trial. In service of this aim, we examined test-retest reliability and practice effects associated with VRFCAT performance in young and older healthy adults and compared these with a standard measure of functional capacity, the UCSD Performance-based Skills Assessment, Validation of Intermediate Measures version (UPSA-2-VIM). Age-related differences in VRFCAT performance were evaluated to assess the sensitivity of the measure to functional declines associated with normal aging. The validity of the VRFCAT was examined by determining correlations between the VRFCAT and cognitive performance domains in young and older healthy adults. Finally, the tolerability of the VRFCAT was evaluated with the use of post-test questionnaires completed by subjects in each age group.

Methods

Participants

Participants included 44 healthy Young Adults (YAs) ages 18-30 (20 female) and 41 healthy Older Adults (OAs) ages 55-70 (24 female). Participants participated as a part of a larger investigation and validation of the VRFCAT instrument in schizophrenia (17). All participants included in the following report were from the control group of that larger validation project. Participants were recruited from three academic sites: the University of California- San Diego, the University of Miami Miller School of Medicine, and the University of South Carolina. Each participant received compensation in the form of pre-paid gift cards in the amount of $50 (Visit 1, 3.5 hours) and $25 (Visit 2, 1.5 hours). Detailed demographic information is displayed in Table 1.

Measures

The Virtual Reality Functional Capacity Assessment Tool (VRFCAT)—The Virtual Reality Functional Capacity Assessment Tool (VRFCAT) is a computer-based virtual-reality measure of functional capacity that relies on a realistic simulated environment to recreate routine IADLs (15, 16). The VRFCAT consists of a tutorial and 6 alternate versions to allow for repeated assessment and evaluation of change over time. The VRFCAT presents participants with multiple instrumental activities of daily living including: navigating a kitchen, catching a bus to a grocery store, finding/purchasing food in a grocery store, and returning home on a bus. Participants complete these scenarios through a progressive storyboard design. See Figure 1 for a screen shot of the kitchen and grocery store scenarios.

University of California, San Diego Performance-Based Skills Assessment (UPSA-2-VIM)—In addition to the VRFCAT, participants completed the UPSA-2-VIM, a standard rater-administered performance-based measure of functional capacity utilizing physical props and materials (18-20). The UPSA-2-VIM takes about 30 minutes to

administer and measures performance in several domains of everyday living, such as counting money, planning an outing, and medication management.

MATRICS Consensus Cognitive Battery (MCCB) (21)—The MCCB is a standardized neurocognitive battery developed as a part of the National Institute of Mental Health (NIMH) Initiative, Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS). The battery includes 10 tests to assess performance in 7 cognitive domains. Tests include: BACS symbol coding, category fluency, Trail Making Test- Part A (Speed of processing domain), Continuous Performance Test – Identical Pairs (CPT-IP, Attention/vigilance), WMS III Spatial Span, Letter Number Span (LNS, Working memory), Hopkins Verbal Learning Test-Revised (HVLT-R - learning trials only, Verbal learning domain), Brief Visuospatial Memory Test- Revised (BVMT-R - learning trials only, visual learning domain), and MSCEIT Managing Emotions (social cognition domain).

Procedure

All participants provided written informed consent prior to engagement in any study-related activities. Each participant completed two study visits. At Visit 1, participants completed a demographic questionnaire and provided detailed information regarding computer use. Next, they completed the VRFCAT, UPSA-2-VIM and MCCB assessments. The administration order of the VRFCAT and UPSA-2-VIM was counterbalanced across participants. Participants completed an alternate version of the VRFCAT and repeated the UPSA-2-VIM at Visit 2 which took place 7-14 days later. At both visits, participants completed a VRFCAT questionnaire that asked them to rate pleasantness, ease of use, clarity of instructions and realism of the VRFCAT virtual environment on a 7-point Likert scale. The duration of study visits was approximately 3.5 hours for Visit 1 and 1.5 hours for Visit 2, inclusive of breaks between tasks.

Analysis

Demographic and baseline characteristics were compared between young and older adults using the Wilcoxon rank sum test or the chi-square test for independence. Results for the VRFCAT summary measures were transformed into T-scores based on data from Visit 1. T-score values were inverted such that higher scores indicated better performance. In order to preserve differences of interest between groups, T-scores for the VRFCAT and MCCB were not corrected for age and gender (22). Pearson correlations for VRFCAT T-scores and the UPSA-2-VIM total score with the MCCB T-scores were calculated separately within each age group. Differences in the correlations between age groups were evaluated using a Fisher's z transformation of the correlation coefficients (23). Practice effects were examined between Visits 1 and 2 for the VRFCAT raw scores and the UPSA-2-VIM total score. Effect size of the practice effect was calculated using Cohen's d (24), and significance of the paired differences was tested using the Wilcoxon signed-rank test. Test-retest reliability of the VRFCAT and UPSA-2-VIM between Visits 1 and 2 was calculated using the intraclass correlation coefficient (ICC).

Results

Demographic characteristics are presented in Table 1 for the 44 young adults, ages 18-30, and the 41 older adults, ages 55-70. In addition to age, the two groups were significantly different with respect to employment and computer skills. Young adults were more likely to be employed either part-time or full-time (68 vs. 29%). The entire group of young adults reported being comfortable using the computer as compared to 90% of the older adults.

Functional capacity, as measured by the VRFCAT and the UPSA-2-VIM, and cognition, as measured by the MCCB composite score, are presented in Table 2 for the two age groups at Visit 1. Older subjects took an average of 3 minutes longer to complete the VRFCAT and made an average of 2 more errors during the test than their younger counterparts. In addition, older adults scored a full standard deviation lower on the unadjusted MCCB composite score compared with young adults. Figure 2 divides the total time for completion of the VRFCAT into that spent on each individual objective. Older adults took significantly longer on all objectives, and were noticeably slower on objectives 3, 9, 10, and 12 (Figure 2). Older adults also made significantly more errors than young subjects on these objectives and on objective 7 (Figure 3).

Correlations between functional capacity measures and cognition are presented in Table 3. The VRFCAT and the UPSA-2-VIM demonstrated strong correlations with the MCCB composite as well as many of the subtests across both age groups. Higher scores on the VRFCAT were associated with better performance on the Trail Making Test A for older adults and the MSCEIT for young adults. The UPSA-2-VIM demonstrated a strong positive correlation with the HVLT for both age groups, but significantly more so for young adults.

Practice effects and test-retest reliability are provided in Table 4 for the VRFCAT measures and UPSA-2-VIM. Differences between Visits 1 and 2 were not statistically significant for any VRFCAT measures. In contrast, a practice effect of 2.7 points (d=0.3) for the UPSA-2-VIM was noted for both older (p=0.018) and young adults (p=0.005). Test-retest reliability was greatest for the VRFCAT Time and Recipe measures, and for the UPSA-2-VIM. Although UPSA-2-VIM reliability was the same for young and older subjects (ICC=0.72), VRFCAT Time demonstrated greater reliability among young adults (ICC=0.80 vs. 0.64). The VRFCAT recipe measure demonstrated better reliability in older adults (ICC=0.77 vs. 0.63), and was the most reliable functional capacity endpoint in the older age group across measures.

As described in Table 5, subjects on average found the VRFCAT to be highly realistic, pleasant to take, easy to use, and found the instructions clear. Although all subjects rated the task highly with respect to ease of use and understandability of instructions (mean ratings were > 6 out of 7 for both age groups), these ratings were significantly higher for subjects in young age group.

Discussion

For an anti-dementia treatment to receive marketing approval for mild to severe Alzheimer disease from the FDA, it must demonstrate efficacy on a cognitive measure and a functional

or global measure. While the requirements for earlier stages of the illness are in a state of ongoing development (25), it is clear that new tools for assessment of clinically meaningful change will be needed (26). Assessment of functional capacity in primary prevention and preclinical/prodromal AD trials will require measures with improved sensitivity to changes in non-demented individuals.

Many previous studies have relied on partner-reported measures that require the availability of a competent informant and may lack sensitivity to subtle functional deficits in this population. Performance- based measures represent a viable alternative to interview-based measures that would not require informant involvement (11). This method may be particularly attractive for screening of healthy individuals or tracking prodromal subjects over time. However, these measures need to demonstrate good test-retest reliability, minimal practice effects, correlations with cognitive measures, correlations with functional outcomes, practicality for clinical investigators and clinicians, and tolerability for subjects in clinical trials.

Data collected in this study suggest that the VRFCAT demonstrates many of the necessary characteristics for a measure of clinically meaningful change in preclinical AD, MCI and AD clinical trials. Test-retest reliability was good, with an intra- class correlation of .80 in young adults and .67 in older adults. Furthermore, the VRFCAT had very small practice effects in contrast to the UPSA-2-VIM, a standard measure of functional capacity, suggesting that it may be very useful in clinical trials with repeated testing.

Performance on the VRFCAT, like the UPSA-2-VIM, was strongly correlated with cognitive performance, with Pearson correlations on cognitive composite scores ranging between .66 and .79 in older and young subject groups, respectively. These results suggest that the VRFCAT meets the important criterion of strong relationships with cognitive performance. All but one of the cognitive tests had correlations with VRFCAT Total time that were greater than r=.30, suggesting that many cognitive domains are involved in VRFCAT performance in young and older subjects alike.

The VRFCAT also demonstrated sensitivity to age-related differences. As expected, older subjects required more time to complete the VRFCAT scenarios and made more errors, especially on the more difficult scenarios. While overall cognition and VRFCAT performance were similarly highly correlated in young and older subjects, VRFCAT total time and errors was significantly more correlated with social cognition in young subjects than in older subjects, suggesting that social cognition may not be a relevant component of performance in older subjects.

It was not appropriate to evaluate practicality from the perspective of the tester in this study since very few testers were involved in the study and all sites were highly active academic research centers with extensive experience with functional capacity measures. However, we were able to assess tolerability from the perspective of the subjects who participated in the study. All of the young subjects and 90% of older subjects indicated a high degree of familiarity and comfort with computers, indicating computerized testing is appropriate in this population. In addition, subjects on average found the VRFCAT to be realistic, pleasant

to take, easy to use, and found the instructions clear. Older subjects had a little more trouble on average with regard to the ease of use and clarity of instructions, with a few older individuals having particular trouble. These results suggest that it may be important to identify those few older subjects in clinical trials who may be befuddled by computerized tasks. Screening participants for computer comfort and literacy is easily accomplished.

In summary, the VRFCAT is a new instrument for the assessment of functional capacity that demonstrates good test-retest reliability, strong correlations with cognitive performance, expected age-related decline, and reported tolerability in most older subjects. These findings provide preliminary support for the VRFCAT as a potential measure of functionally relevant change in clinical trials for primary prevention, preclinical AD and MCI.

Acknowledgments

References

1. McKhann GM, Knopman DS, Chertkow H, 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; 7:263–269. [PubMed: 21514250]

2. Marshall GA, Amariglio RE, Sperling RA, Rentz DM. Activities of daily living: Where do they fit in the diagnosis of Alzheimer's disease. Neurodegener Dis Manag. 2012; 2:483–491. [PubMed: 23585777]

3. Albert MS, DeKosky ST, Dickson D, 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; 7:270–279. [PubMed: 21514249]

4. Amieva H, Le Goff M, Millet X, et al. Prodromal Alzheimer's disease: Successive emergence of the clinical symptoms. Ann Neurol. 2008; 64:492–498. [PubMed: 19067364]

5. Sikkes SA, De Lange-De Klerk ES, Pijnenburg YA, Scheltens P, Uitdehaag BM. A systematic review of instrumental activities of daily living scales in dementia: room for improvement. J Neurol Neurosurg Psychiatry. 2009; 80:7–12. [PubMed: 19091706]

6. Galasko D, Bennett D, Sano M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Disord. 1997; 11:S33–S39. [PubMed: 9236950]

7. Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol. 1982; 37:323–329. [PubMed: 7069156]

8. Conde-Sala JL, Reñe-Ramírez R, Turro-Garriga O, et al. Factors associated with the variability in caregiver assessments of the capacities of the patients with Alzheimer's disease. J Geriatr Psychiatry Neurol. 2013; 26:86–94. [PubMed: 23514974]

9. Arguelles S, Loewenstein DA, Eisdorfer C, Argüelles T. Caregivers' judgments of the functional abilities of the Alzheimer's disease patient: impact of caregivers' depression and perceived burden. J Geriatr Psychiatry Neurol. 2001; 14:91–98. [PubMed: 11419574]

10. Dassel KB, Schmitt FA. The impact of caregiver executive skills on reports of patient functioning. Gerontologist. 2008; 48:781–792. [PubMed: 19139251]

11. Gomar JJ, Harvey PD, Bobes-Bascaran MT, Davies P, Goldberg TE. Development and cross-validation of the UPSA short form for the performance-based functional assessment of patients with mild cognitive impairment and Alzheimer disease. Am J Geriatr Psychiatry. 2011; 19:915–922. [PubMed: 22024615]

12. Arguelles S, Loewenstein DA, Eisdorfer C, Argüelles T. Caregivers' judgments of the functional abilities of the Alzheimer's disease patient: impact of caregivers' depression and perceived burden. J Geriatr Psychiatry Neurol. 2001; 14:91–98. [PubMed: 11419574]

13. Green MF, Nuechterlein KH, Kern RS, et al. Functional co-primary measures for clinical trials in schizophrenia: Results from the MATRICS psychometric and standardization study. Am J Psychiatry. 2008; 165:221–228. [PubMed: 18172017]

14. Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol. 1982; 37:323–329. [PubMed: 7069156]

15. Ruse SA, Davis VG, Atkins AS, et al. Development of virtual reality assessment of everyday living skills. J Vis Exp. 2014; 86:1–8.

16. Ruse SA, Harvey PD, Davis VG, Atkins AS, Fox, Keefe RS. Virtual Reality Functional Capacity Assessment in Schizophrenia: Preliminary data regarding feasibility and correlations with cognitive and functional capacity performance. Schizophr Res Cogn. 2014; 1:21–26.

17. Keefe, RSE.; Ruse, S.; Hufford, et al. Virtual Reality Functional Capacity Assessment: Development and Validation of a Computerized Assessment of Functional Skills; Symposium abstract, 13th International Congress on Schizophrenia Research meeting (ICOSR); Orlando, FL. 4/20/2013–4/25/2013;

18. Patterson TL, Goldman S, McKibbin CL, et al. UCSD performance based skills assessment: Development of a new measure of everyday functioning for severely mentally ill adults. Schizophr Bull. 2001; 27:235–245. [PubMed: 11354591]

19. Goldberg TE, Koppel J, Keehlisen L, et al. Performance-based measures of everyday function in mild cognitive impairment. Am J Psychiatry. 2010; 167:845–853. [PubMed: 20360320]

20. Green MF, Schooler NR, Kern RS, et al. Evaluation of functionally-meaningful measures for clinical trials of cognition enhancement in schizophrenia. Am J Psychiatry. 2011; 168:400–407. [PubMed: 21285142]

21. Nuechterlein, KH.; Green, MF. MATRICS Consensus Cognitive Battery Manual. MATRICS Assessment, Inc; Los Angeles, CA: 2006.

22. Nuechterlein KH, Green MF, Kern RS, et al. The MATRICS consensus cognitive battery; Part 1: Test selection, reliability, and validity. Am J Psychiatry. 2008; 165:203–213. [PubMed: 18172019]

23. Cohen, J.; Cohen, P.; West, SG.; Aiken, LS. Applied multiple regression/correlation analysis for the behavioral sciences. 3rd. Taylor and Francis Group LLC; 2003.

24. Cohen, J. Statistical power analysis for the behavioral sciences. New Jersey: Lawrence Erlbaum Associates; 1988.

25. Kozauer N, Katz R. Regulatory innovation and drug development for early-stage Alzheimer's disease. N Eng J Med. 2013; 368:1169–1171.

26. Keefe, RSE. Challenges with Defining Clinical Meaningfulness in CNS Diseases; Workshop at the ACT-AD FDA/AD Allies meeting; 11/5/2014–11/6/2014;

A	VRFCAT Objective #	Objective
	1	Move your mouse over the recipe on the counter. Click the left mouse button, highlight the 'pick up' action, and click the left mouse button again.
	2	Look for ingredients in your cabinets and refrigerator. When highlighted, click the left mouse button, highlight the 'open' action and click the left mouse button again. (see Screen Shot)
	3	Access your recipe and cross off the ingredients that you already have in your apartment. When you are finished crossing off the ingredients pick up the bus schedule on the counter.
	4	Pick up the billfold on the counter. When you have the billfold highlighted, click the left mouse button and highlight the 'pick up' action and click the left mouse button.
	5	Leave the apartment and head to the bus stop. When you have the apartment door highlighted, click the left mouse button and highlight the 'leave apartment' action, then click the left mouse button.
	6	Wait for the correct bus to the grocery store and then board it when it arrives.
	7	Add up the exact amount of bus fare in your hand. Once you have the exact change for the bus fare in your hand, highlight the confirmation button with your mouse cursor and left click.
	8	Select a food aisle to begin shopping. Move your mouse cursor over the aisle button that you want to go to. When the button is highlighted, press the left mouse button. (See Screen Shot)
	9	Continue shopping for the necessary food ingredients. When you have all of the necessary ingredients in your shopping cart, click the 'check out' button.
	10	Add up the exact amount for your purchase in your hand. Once you have the exact change for the purchase in your hand, highlight the confirmation button with your mouse cursor and left click.
	11	Wait for the same bus that took you to the grocery store and then board it when it arrives.
	12	Add up the exact amount of bus fare in your hand. Once you have the exact change for the bus fare in your hand, highlight the confirmation button with your mouse cursor and left click.

Figure 1. Virtual Reality Functional Assessment Tool (VRFCAT)

The VRFCAT allows for assessment of completion time and errors within and across 12 performance objectives described here (panel A). Screenshots (panel B) provide examples of the virtual environment as the subject performs tasks in the kitchen at home (objectives 1-5), at the bus stop (objectives 6-7 and 11-12), and in the grocery store (objectives 8-10).
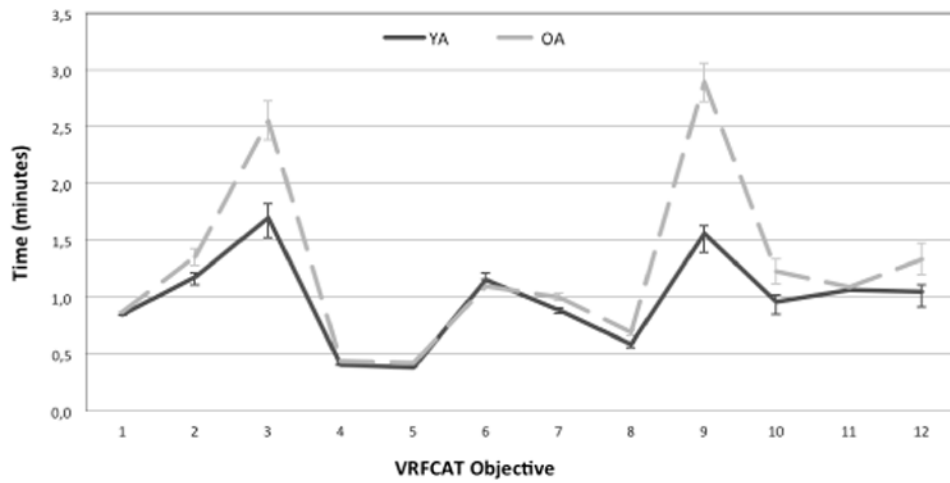
Figure 2. Mean Completion Time on VRFCAT Objectives for Young and Older Adults
Mean completion time for each objective is displayed in minutes. Error bars represent +/-
one standard error. Older adults took significantly longer than younger subjects on all
VRFCAT objectives; age differences in completion time were largest for objectives 3, 9, 10,
and 12.

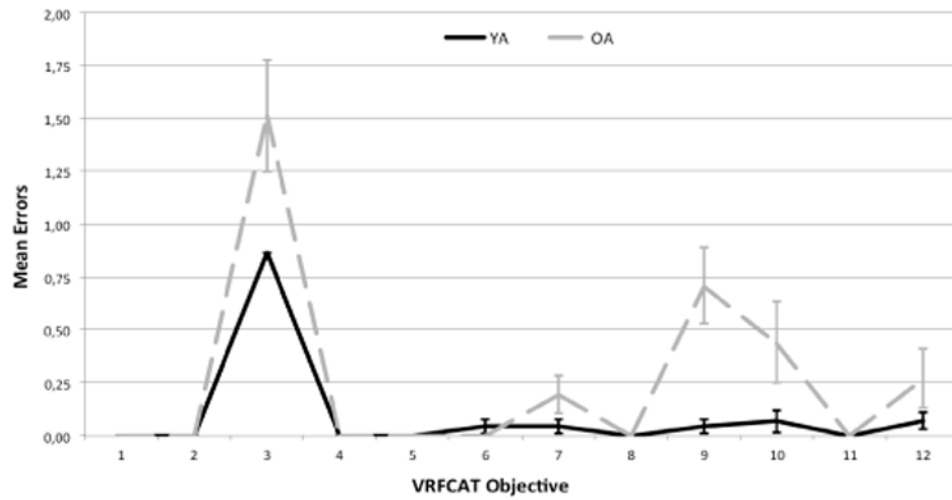**Figure 3. Mean Number of Errors on VRFCAT Objectives for Young and Older Adults**
Mean number of errors displayed for both young and older adults on each objective. Error bars represent +/- one standard error. Older adults made significantly more errors than younger subjects on objectives 3, 7, 9, 10, and 12.

Table 1

Sample Demographics

	YA 18-30 yo (N=44)	OA 55-70 yo (N=41)	p-value
Age, Mean (St Dev)	25.8 (3.47)	60.8 (4.38)	< 0.001
Male, N (%)	24 (55)	17 (41)	0.224
Caucasian, N (%)	25 (57)	23 (56)	0.947
Years of Education, Mean (St Dev)	14.8 (2.28)	14.9 (2.95)	0.873
Employed, N (%)	30 (68)	12 (29)	< 0.001
Comfortable with PC, N (%)	44 (100)	37 (90)	0.035

Demographic information, employment status, and mean level of comfort with computers for participants in the young and older age groups. P-values for group comparisons of means are based on the Wilcoxon two-sample rank sum analysis; p-values for comparisons of frequency counts are based on the chi-square test for independence.

Table 2

Functional capacity and cognitive performance by age

	YA (N=44)	OA (N=41)	p-value	Cohen's d
VRFCAT Summary Measures	**Mean ± SD**	**Mean ± SD**		
Total Time (minutes)	11.8 ± 2.09	15.0 ± 3.28	< 0.001	1.2
Total Errors	1.1 ± 1.50	3.1 ± 3.35	< 0.001	0.8
Total Progressions	0.1 ± 0.21	0.5 ± 0.81	< 0.001	0.9
Total Bus Schedule Checks	3.5 ± 1.90	3.7 ± 1.73	0.463	0.1
Total Recipe Checks	12.4 ± 4.81	11.8 ± 4.99	0.475	-0.1
UPSA-2-VIM Total Score	84.4 ± 8.63	83.1 ± 8.88	0.562	0.1
MCCB Composite T-Score (uncorrected)	49.0 ± 11.67	36.9 ± 12.00	< 0.001	1.0

Age differences in functional capacity (VRFCAT, UPSA-2-VIM) and cognitive performance (MCCB) at Visit 1. For all VRFCAT summary measures, a higher score indicates worse performance. P-values for group comparisons of means are based on the Wilcoxon two-sample rank sum analysis. Positive values of Cohen's d indicate poorer performance in OAs relative to YAs.

Table 3

Correlations between functional capacity and cognitive performance in young and older adults

Young Adults (18-30 yo)	MCCB Composite	TMT	BACSSC	HVLT	WMSIII	LNS	NAB	BVMT	Fluency	MSCEIT	CPT
VRFCAT Time	0.79	0.19	0.35	0.51	0.50	0.50	0.47	0.44	0.49	0.72*	0.63
VRFCAT Errors	0.67	0.09*	0.40	0.50	0.38	0.46	0.35	0.45	0.25	0.59*	0.47
VRFCAT Progressions	0.37	0.12	0.21	0.26	0.21	0.24	0.08	0.28	0.18	0.33	0.37
VRFCAT Bus Schedule Checks	0.20	0.11	-0.12	0.16	0.06	0.30	0.07	0.33	0.18	0.10	0.09
VRFCAT Recipe Checks	0.33	-0.08	-0.10	0.36	0.16	0.36	0.06	0.26	0.36	0.33	0.17
UPSA-2-VIM	0.79	0.17	0.39	0.66*	0.46	0.59	0.48	0.58	0.49	0.53	0.54

Older Adults (55-70 yo)	MCCB Composite	TMT	BACSSC	HVLT	WMSIII	LNS	NAB	BVMT	Fluency	MSCEIT	CPT
VRFCAT Time	0.66	0.48	0.47	0.49	0.47	0.62	0.50	0.47	0.35	0.32*	0.45
VRFCAT Errors	0.55	0.45*	0.45	0.39	0.53	0.59	0.43	0.46	0.22	0.10*	0.34
VRFCAT Progressions	0.37	0.28	0.23	0.41	0.16	0.47	0.25	0.33	0.21	0.03	0.22
VRFCAT Bus Schedule Checks	0.41	0.09	0.12	0.41	0.34	0.48	0.28	0.32	0.44	0.03	0.29
VRFCAT Recipe Checks	0.05	-0.08	-0.07	0.24	-0.09	0.27	-0.06	0.03	0.13	-0.06	-0.01
UPSA-2-VIM	0.73	0.50	0.61	0.43*	0.64	0.57	0.43	0.58	0.30	0.41	0.60

Correlations were computed using T-scores for VRFCAT and MCCB measures. T-scores were not corrected for age or gender. MCCB Subtests include: Trail Making Test, Part A (TMT); Brief Assessment of Cognition Symbol Coding (BACSSC); Hopkins Verbal Learning Test-Revised (HVLT) ; Wechsler Memory Scale-III - Spatial Span (WMS); Letter Number Span (LNS); Neuropsychological Assessment Battery Mazes (NAB); Brief Visuospatial Memory Test – Revised Learning Trials (BVMT); Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT); Continuous Performance test-Identical Pairs (CPT). Shading identifies the correlations demonstrating significant differences between age groups (p<.05).

Table 4

Practice Effects and Test-Retest Reliability for the VRFCAT and UPSA-2-VIM by Age Group

Assessment	Visit 1 Mean (SD)		Visit 2 Mean (SD)		Difference Mean (SD)		Cohen's d[1]		Intraclass Correlation Coefficient (ICC)	
	YA	OA	YA	OA	YA	OA	YA	OA	YA	OA
VRFCAT Total Time (minutes)	11.8 (2.10)	14.6 (2.52)	11.5 (2.25)	14.3 (3.45)	0.3 (1.38)	0.3 (2.56)	0.1	0.1	0.80	0.64
VRFCAT Total Errors	1.1 (1.46)	2.8 (3.04)	0.9 (1.28)	2.8 (4.65)	0.2 (1.43)	0.0 (4.19)	0.1	0.0	0.46	0.44
VRFCAT Total Progressions	0.0 (0.22)	0.5 (0.72)	0.0 (0.22)	0.4 (0.93)	0.0 (0.22)	0.1 (0.82)	0.0	0.1	0.48	0.52
Total Bus Schedule Checks	3.4 (1.91)	3.5 (1.67)	3.1 (1.70)	3.8 (2.39)	0.3 (2.00)	-0.3 (2.49)	0.2	-0.1	0.39	0.28
Total Recipe Checks	12.5 (4.91)	11.9 (5.10)	12.2 (5.46)	11.9 (5.23)	0.3 (4.50)	0.0 (3.55)	0.1	0.0	0.63	0.77
UPSA-2-VIM	84.8 (8.45)	83.4 (8.98)	87.6 (8.21)	86.2 (9.56)	2.7 (5.76)**	2.7 (6.58)*	0.3	0.3	0.72	0.72

Positive differences and Cohen's d values indicate improvement from Visit 1 to Visit 2. Practice effects for all VRFCAT measures were small and insignificant in both age groups. A practice effect of 2.7 points (d=0.3) for the UPSA-2-VIM was observed in both older and younger adults (p=0.018 and p=0.005, respectively).

Table 5

Mean VRFCAT tolerability ratings by age group

	YA (N=44)	OA (N=41)	p-value[1]
Pleasantness, Mean (St Dev)	5.7 (1.47)	5.9 (1.36)	0.501
Ease of Use, Mean (St Dev)	6.8 (0.64)	6.1 (1.53)	0.004
Instructions, Mean (St Dev)	6.8 (0.70)	6.2 (1.41)	0.006
Realistic, Mean (St Dev)	6.0 (1.25)	6.1 (1.48)	0.468

Subject tolerability measures ranged from 1-7 with higher scores indicating higher levels of tolerability. P-values for comparison of means are based on the Wilcoxon two-sample rank sum analysis.