Self-Administered Cognitive Testing by Older Adults At-Risk for Cognitive Decline

E. Tsoy[1], **K.L. Possin**[1,2], **N. Thompson**[3], **K. Patel**[4], **S.K. Garrigues**[4], **I. Maravilla**[4], **S.J. Erlhoff**[1], **C.S. Ritchie**[2,5]

[1.]Memory and Aging Center, Department of Neurology, University of California, San Francisco, San Francisco, California, USA

[2.]Global Brain Health Institute, University of California, San Francisco, San Francisco, CA, USA

[3.]Osher Center for Integrative Medicine, University of California, San Francisco, San Francisco, California, USA

[4.]Division of Geriatrics, Department of Medicine, University of California, San Francisco, San Francisco, California, USA

[5.]Division of Palliative Care and Geriatric Medicine, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

Abstract

Self-administered computerized cognitive testing could effectively monitor older individuals at-risk for cognitive decline at home. In this study, we tested the feasibility and reliability of 3 tablet-based executive functioning measures and an executive composite score in a sample of 30 older adults (age 80±6) with high multimorbidity. The tests were examiner-administered at baseline and then self-administered by the participants at home across 2 subsequent days. Eight of the participants reported no prior experience with touchscreen technology. Twenty-seven participants completed both self-administered assessments, and 28 completed at least one. Cronbach's alpha (individual tests: .87-.89, composite: .93) and correlations between examiner-administered and self-administered performances (individual tests: .72-.91, composite: .93) were high. The participants who had never used a smartphone or a tablet computer showed comparable consistency. Remote self-administered tablet-based testing in older adults at-risk for cognitive decline is feasible and reliable, even among participants without prior technology experience.

Keywords

Corresponding Author: Elena Tsoy, PhD, Memory and Aging Center, Department of Neurology, University of California San Francisco, 675 Nelson Rising Lane, Suite 190, San Francisco, California, USA, 94158, Tel: 415-514-7477, elena.tsoy@ucsf.edu.

Introduction

Accurate and reliable measurement of cognitive abilities is important for detecting cognitive disorders, evaluating the effects of medications and other interventions on brain health, and optimizing brain care. Self-administered cognitive assessments at home could reduce costs and patient burden and could allow for more frequent measurements for monitoring the dynamics of cognitive functioning overtime (1). Computerized assessments, with opportunities for automated stimuli presentation, scoring and data upload, is the optimal framework for unsupervised cognitive testing. Older adults are using computerized devices with increasing frequency with 68% reporting use of smartphones and 52% reporting use of tablets (2), and may find touchscreen devices particularly comfortable and intuitive to use (3).

To date, only a few studies have reported findings on the feasibility, reliability, or validity of self-administered cognitive assessments in home settings. For example, Rentz et al. (4) found high reliability and in-home feasibility of a brief tablet-based cognitive battery in 49 cognitively normal older adults over 5 sessions across 1 week (57% completed 5/5 and 98% completed 4/5 sessions correctly). Ruano et al. (5) reported 87.6% completion rates and moderate-to-high test-retest reliability of a brief web-based cognitive battery over 3 sessions across 6 months in a sample of 129 non-demented older adults. Finally, Jongstra et al. (6) used a brief smartphone-based cognitive battery over 4 sessions across 6 months in 151 older adults and reported mean completions rates of 60% for all 4 sessions and 95% for at least 1 session, and moderate correlations with conventional neuropsychological tests. These studies provide preliminary evidence supporting the feasibility and validity of unsupervised technology-based cognitive assessments among older adults in home settings, but gaps in research remain, including whether old adults at the highest risk for cognitive decline due to multimorbidity, including those with limited experience with computers, can self-administer tests reliably. In fact, the National Institute of Neurological Disorders and Stroke 2019 Alzheimer's Disease and Related Disorders Summit prioritized the need for more research on the use of self-administered cognitive assessments for the detection of cognitive impairment among older adults (7).

Tests of executive functions and speed offer particular value to research and clinical questions when high sensitivity to impairment and to cognitive change is needed. Executive impairment is often an early marker of neurodegenerative disease, including Alzheimer's disease, and is also impacted by common conditions that impact cognition like medication side effects, vascular disease, and sleep disorders (8). Many executive function tests have strong psychometric properties because they are amenable to collecting many responses in a short period of time and to designs that minimize floor and ceiling effects. On a computerized platform, precise stimuli presentation and accuracy of reaction time measurement add to the reliability of measurement (1,3). The purpose of this study was to examine the feasibility and reliability of a novel tablet-based and self-administered executive functioning assessment in older adults at-risk for cognitive decline due to high multimorbidity.

Methods

Participants

This study was approved by the University of California, San Francisco (UCSF) Committee on Human Research and all followed procedures were in accordance with the Helsinki Declaration. Participants were recruited from an outpatient palliative care clinic and a geriatric clinic at UCSF. All participants provided written informed consent. For inclusion, participants were age 70+, English-speaking, and had 2+ chronic medical diagnoses. The chronic conditions were identified via medical record review of ICD-9 and -10 codes and included kidney disease, congestive heart failure, coronary pulmonary disease, diabetes, hypertension, heart disease, cancer, liver disease, depression, and arthritis. Participants who endorsed a chronic condition diagnosis not listed in their medical record were included if they were taking a medication for that condition.

Participants were excluded if they had a cognitive disorder diagnosis as indicated through medical record review or direct consultation with their primary or geriatric care provider at the referring site. In addition, potential participants were asked if they had experienced memory loss over the past year, and if they endorsed this question, they were administered the Mini-Cog (9) and excluded if they scored less than 3/3 points. Of the 30 participants who met these criteria, 12 endorsed memory loss over the past year but passed the Mini-Cog. The Mini-Cog has adequate performance characteristics for detecting dementia but not mild cognitive impairment (9), and so it is likely - given the high rate of endorsed memory loss in the sample - that some participants had undiagnosed mild cognitive impairment. Participants were also excluded if they had visual or motor impairment severe enough to compromise their ability to use an iPad.

Measures and procedures

Participants completed the Lawton Instrumental Activities of Daily Living Scale (IADLs; 10) and the Katz Index of Independence in Activities of Daily Living (ADLs; 11) to assess daily functional abilities. Symptom burden was assessed by using 19 questions selected from the Memorial Symptom Assessment Scale (12).

Cognitive assessment included tests of executive functions and processing speed, Match (13), cognitive inhibition, Flanker (14), and spatial working memory (Stargazer) administered on the TabCAT software platform (memory.ucsf.edu/tabcat) using a 9.7" iPad. See Figure 1 for screenshots and task descriptions. Before testing, to assess familiarity with touchscreen technology, participants were asked how often they used a smartphone and a tablet with response options being "never,""once in a while," "some of the time,""most of the time," or "always."

Participants completed a 15-minute baseline training and assessment session administered by a research coordinator either in clinic (N = 20) or at home (N = 10), per participant preference. During these sessions, they were trained on the basic use of the iPad including how to turn on and charge the device and how to open the testing application. Participants then took the tests with the examiner present, who was available to answer any questions about the test instructions (examiner-administered session). The subsequent 2 testing

sessions were self-administered at home in the next 2 days. After completion of the study, participants completed a survey on whether they found any tests confusing or unclear and, if they endorsed the question, were asked to describe what aspects of the test they found confusing.

Statistical analyses

The raw scores on individual tests were standardized to z-scores based on the mean and standard deviation (SD) of the baseline sample performance to facilitate comparisons between tests and across administration types. The Executive Composite for each testing session was generated by averaging the z-scores on 3 individual tests for respective sessions. To assess overall self-administered performance, we averaged the z-scores of individual tests and the Executive Composite scores across the 2 self-administered sessions.

To evaluate test-retest reliability across examiner- and self-administered sessions, first, Pearson correlations were performed between examiner-administered performance and mean self-administered performance. Next, Cronbach's alpha was used to compare performance across all 3 sessions. We evaluated potential effects of familiarity with touchscreen devices on test-retest reliability by assigning participants into low and high touchscreen familiarity subgroups and then computing Cronbach's alphas separately for these groups.

Additionally, we explored potential practice effects across testing sessions by comparing performances on the Executive Composite via repeated measures ANOVA with post-hoc Bonferroni comparisons. In these analyses, the session number was the independent variable. To explore whether touchscreen familiarity impacted practice effects, we reran the model including the interaction term of session number and touchscreen familiarity (low and high).

All analyses were conducted using SPSS statistical package version 25 (SPSS Inc, Chicago, IL). Tests were two-tailed for all analyses with a significance level set at $p < .05$.

Results

Sample characteristics

Thirty older adults (age: 80±6, gender: 53% female, ethnicity: 90% non-Hispanic White) met inclusion criteria and participated in the study. Eight participants completed some college, 11 had a bachelor's degree, and 11 had a master's degree. Average number of comorbidities in the entire sample was 3.4±1.6. Mean self-reported number of medications taken was 8.9±2.4, although this may be an underestimation as the data collection form for medications allowed up to a maximum of 12 medications only. The mean reported symptom burden score was 6.7±3.5 with 21 (70%) of participants endorsing 5 or more symptoms. Specifically, more than half of the sample endorsed chronic pain (20; 67%), lack of enegery (16; 53%), feeling drowsy (17; 57%), numbness/tingling in hands/feet (17; 57%), and worrying (17; 57%). With regard to daily functioning, 6 participants (20%) endorsed IADLs dependency and 8 participants (27%) endorsed basic ADLs dependency.

Feasibility

Of the 30 participants who completed the baseline assessment, 27 completed both at-home assessments (90%). The reasons for non-completion were as follows: 2 participants experienced technical difficulties launching the cognitive tests and another participant completed only the first self-administered session and did not provide a reason for non-completion of the second.

Test-retest reliability

Correlations between examiner-administered and mean self-administered assessments were highest for the executive composite ($r = .92$), followed by Match ($r = .91$), Stargazer ($r = .79$), and Flanker ($r = .72$; Figure 2). Similarly, Cronbach's alpha across all 3 sessions was highest for the composite ($a = .93$) followed by Match ($a = .89$), Stargazer ($a = .89$) and Flanker ($a = .87$).

Familiarity with touchscreen technology

Of the 27 participants with complete test data, 9 reported "never" using a smartphone, 2 reported using it "most of the time," and 16 reported using one "always." Tablet use was more variable: 9 reported "never" using one, 8 "once in a while," 3 "some of the time," 1 "most of the time," and 6 "always." Of the 9 participants who reported never using a smart phone, 8 also reported "never" using a tablet; the remaining one reported "always" using one. Thus, there were 8 participants who reported that they have never used a touchscreen device (smartphone or tablet), and we categorized them into the "low familiarity" group. The other 19 subjects, who reported that they used either a smart phone or a tablet most of the time or always, were categorized into the "high familiarity group." The Cronbach's alpha analyses were the performed separately for the low and high touchscreen familiarity groups. The values were very similar for the Executive Composite (low $a = .93$, high $a = .93$), Match (low $a = .88$, high $a = .89$), and Flanker (low $a = .91$, high $a = .89$), but were lower in the low familiarity group for Stargazer (low $a = .83$, high $a = .94$).

Familiarity with touchscreen technology was also examined among non-completers. Of the 3 participants with incomplete test data, only 1 had low familiarity.

Practice effects on the Executive Composite

Mean Executive Composite raw scores across the 3 sessions were 0.00 (SD = 0.88), 0.20 (SD = 0.84), and 0.31 (SD = 0.89). Results from a repeated measures ANOVA indicated presence of practice effects ($F = 5.07$, $P = .010$), with a significant mean difference between the first and third sessions of 0.31 ($P = .008$). When the touchscreen familiarity variable was added to the model, only the effect of the session number was significant ($F = 5.42$, $P = .007$; familiarity: $F = 2.48$, $P = .403$; interaction: $F = 0.93$, $P = .403$).

Post-assessment survey

The post-assessment survey was completed by 29 participants. Twenty-two of them (76%) reported that none of the cognitive tests were confusing or unclear, and 7 (24%) reported that one or more of the cognitive tests were confusing or unclear. Of these, 5 identified Stargazer

as the test that caused confusion stating, "[it was] frustrating, the stars were hard and [I] just said to hell with it," and "flashing test and circles too fast and I was unclear about tapping the circles." None of the participants reported either Flanker or Match to be confusing or unclear.

Discussion

We found high reliability and feasibility of self-administered assessment of executive functions by older adults with multimorbidity. Specifically, our findings showed high test-retest reliability of all 3 individual measures and the executive composite across examiner-administered and self-administered sessions. These results compliment widely discussed advantages of computerized cognitive measures, including enhanced precision, and minimization of examiner bias and staffing costs (1). Reliability and completion rates were similar among participants with high and low familiarity with touchscreen devices, which supports the notion of usability of unsupervised computerized testing in older adults regardless of prior experience with this technology. Taken together, our findings add to the body of literature on potential clinical applications of reliable measurement and monitoring of cognitive functioning in older adults at home (4–6), although a number of relevant considerations discussed below warrant attention.

First, test characteristics appear to play an important role in adherence and consistency of performance across settings. In particular, we found highest reliability and consistency estimates for Match, which could be related to the fact that this measure relies on widely distributed brain networks, engages several cognitive processes, and collects many responses in a short period of time (13) while also having brief and easy-to-understand instructions. In contrast, Stargazer had lower consistency among participants with low touchscreen familiarity. This task has the most complex instructions and was described by a few participants as confusing or unclear. This highlights the importance of simple, user-friendly interfaces and instructions for unsupervised at-home applications of cognitive testing in older adults.

Second, we observed practice effects, which is not surprising given the short time between testing sessions and is consistent with prior literature on learning effects on executive functioning tests (5, 6). This finding is an important reminder that correction for practice effects – for example, via a control group – is critical to make valid interpretations of cognitive change.

Third, a number of concerns have been raised regarding technological aspects of computerized cognitive tests (1). Specifically, the wide variability in hardware and software platforms may impact reliability and validity of cognitive findings. Our approach minimizes these concerns by using managed devices, thus ensuring homogeneity of stimuli presentation, response time measurement, and other logistical issues that may arise by using different data collection devices. Additionally, 2 participants in our study experienced technical difficulties resulting in incomplete data. A well-thought-out user interface, including an easy to launch application and the availability of technical support and guidance, may minimize missing data due to technical issues.

Conclusion

The potential value of self-administered computerized cognitive assessments for clinical and research applications is being increasingly recognized (1, 15). The present results demonstrate that consistent cognitive scores can be obtained among older adults with multimorbidity when self-testing at home, and that consistency is similar among those with and without prior experience with touchscreen computing devices. Software programs that are easy to launch and use, and cognitive tests that are simple to understand, may maximize adherence and reliability. They also may enhance our ability to understand the day-to-day impact of an array of environmental and biological events on cognitive function.

Acknowledgements:

References

1. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. Clin Neuropsychol. 2012;26(2):177–196. doi:10.1080/13854046.2012.663001 [PubMed: 22394228]

2. Vogels EA; Pew Research Center. Millennials stand out for their technology use, but older generations also embrace digital life. https://www.pewresearch.org/fact-tank/2019/09/09/us-generations-technology-use/ Published September 9, 2019. Accessed December 20, 2019.

3. Canini M, Battista P, Della Rosa PA, et al. Computerized neuropsychological assessment in aging: testing efficacy and clinical ecology of different interfaces. Comput Math Methods Med. 2014;2014:804723. doi:10.1155/2014/804723 [PubMed: 25147578]

4. Rentz DM, Dekhtyar M, Sherman J, et al. The feasibility of at-home iPad cognitive testing for use in clinical trials. J Prev Alzheimers Dis. 2016;3(1):8–12. doi:10.14283/jpad.2015.78 [PubMed: 26998469]

5. Ruano L, Sousa A, Severo M, et al. Development of a self-administered web-based test for longitudinal cognitive assessment. Sci Rep. 2016;6:19114. doi:10.1038/srep19114 [PubMed: 26743329]

6. Jongstra S, Wijsman LW, Cachucho R, Hoevenaar-Blom MP, Mooijaart SP, Richard E. Cognitive testing in people at increased risk of dementia using a smartphone app: the iVitality proof-of-principle study. JMIR Mhealth Uhealth. 2017;5(5):e68. doi:10.2196/mhealth.6939 [PubMed: 28546139]

7. Schneider J, Jeon S, Gladman JT, Corriveau RA. ADRD summit 2019 report to the National Advisory Neurological Disorders and Stroke Council. https://www.ninds.nih.gov/sites/default/files/2019_adrd_summit_recommendations_508c.pdf Published 9 4, 2019. Accessed December 20, 2019.

8. Rabinovici GD, Stephens ML, Possin KL. Executive dysfunction. Continuum (Minneap Minn). 2015;21:646–659. doi:10.1212/01.CON.0000466658.05156.54 [PubMed: 26039846]

9. Holsinger T, Plassman BL, Stechuchak KM, Burke JR, Coffman CJ, Williams JW Jr. Screening for cognitive impairment: comparing the performance of four instruments in primary care. J Am Geriatr Soc. 2012;60(6):1027–1036. doi:10.1111/j.1532-5415.2012.03967.x [PubMed: 22646750]

10. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. 1969;9:179–186. doi: 10.1093/geront/9.3_Part_1.179 [PubMed: 5349366]

11. Katz S, Downs TD, Cash HR, Grotz RC. Progress in development of the index of ADL. Gerontologist. 1970;10:20–30. doi:10.1093/geront/10.1_part_1.20 [PubMed: 5420677]

12. Portenoy RK, Thaler HT, Kornblith AB, et al. The Memorial Symptom Assessment Scale: an instrument for the evaluation of symptom prevalence, characteristics and distress. Eur J Cancer. 1994;30A:1326–1336. doi: 10.1016/0959-8049(94)90182-1 [PubMed: 7999421]

13. Possin KL, Moskowitz T, Erlhoff SJ, et al. The Brain Health Assessment for detecting and diagnosing neurocognitive disorders. J Am Geriatr Soc. 2018;66(1):150–156. doi:10.1111/jgs.15208 [PubMed: 29355911]

14. Kramer JH, Mungas D, Possin KL, et al. NIH EXAMINER: conceptualization and development of an executive function battery. J Int Neuropsychol Soc. 2014;20(1):11–19. doi:10.1017/S1355617713001094 [PubMed: 24103232]

15. Dodge HH, Zhu J, Mattek NC, Austin D, Kornfeld J, Kaye JA. Use of high-frequency in-home monitoring data may reduce sample sizes needed in clinical trials. PLoS One. 2015;10(9):e0138095. doi:10.1371/journal.pone.0138095 [PubMed: 26379170]
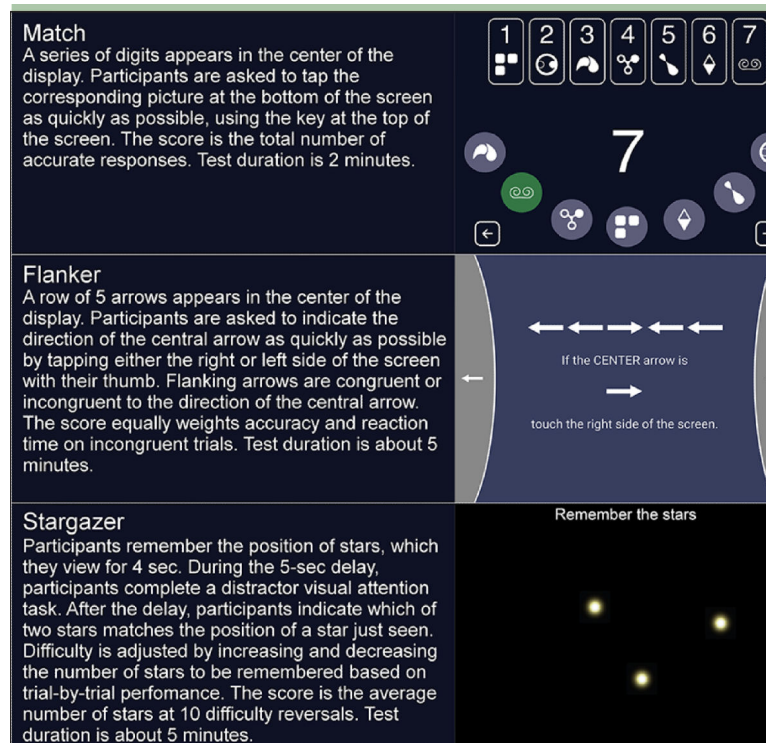
Figure 1.
Task descriptions
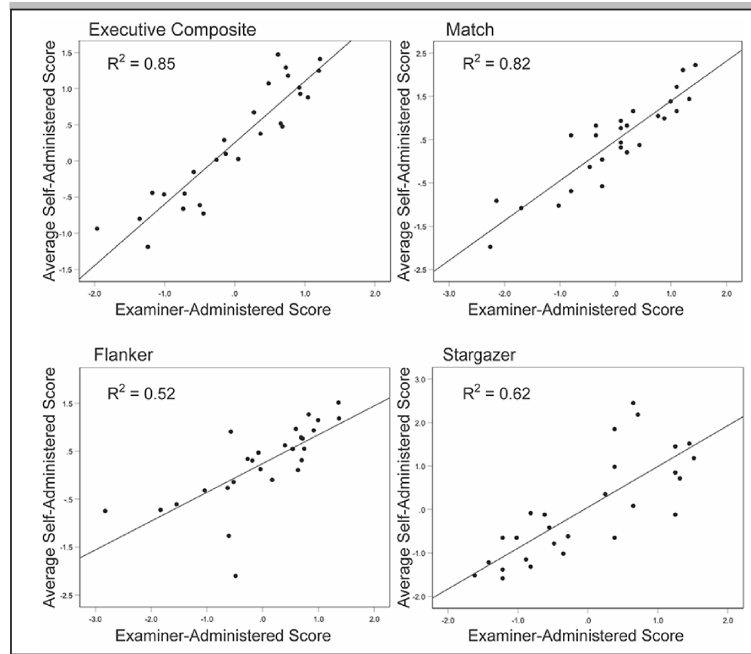
Figure 2.
Correlations between examiner-administered and self-administered session scores