

Feasibility of remote unsupervised cognitive screening with SATURN in older adults

Chiara F. Tagliabue^a, David Bissig^b, Jeffrey Kaye^c, Veronica Mazza^a, Sara Asseconi^{a*}

^a Center for Mind/Brain Sciences - CIMEC, University of Trento, Rovereto (TN), Italy, 38068

^b Department of Neurology, University of California–Davis, Sacramento, California, USA, CA 95616

^c Department of Neurology, Oregon Health and Science University, Portland, Oregon, USA, OR 97239

Running head: "SATURN: remote cognitive screening"

** Corresponding author*

Sara Asseconi,

Centre for Mind/Brain Sciences - CIMEC,

University of Trento,

Corso Bettini 31, Rovereto (TN), 38068, Italy.

Email: Sara.asseconi@gmail.com; sara.asseconi@unint.it

Acknowledgements: The authors thank Greta Varesio and Giulia Buzi for their help with data collection and literature review.

Declaration of conflicting interests: S.A is named inventor on a pending patent, submitted by the University of Birmingham (publication number WO/2022/106850). C.F.T, D.B., J.K., and V.M. declare that they have no conflict of interest.

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Fondazione Cassa di Risparmio di Trento e Rovereto (CARITRO) [grant number 000040103444CARITRO]. JAK is supported by National Institute on Aging (NIA) [grant numbers P30 AG066518 and P30 AG024978]. The sponsors were not involved in the study design or in the collection, analysis, and interpretation of data.

Ethics approval: The study was approved by the University of Trento Research Ethics Committee (Protocol No. 2021-041).

Abstract

Widespread cognitive test screening as part of tele-public health initiatives necessitates a test that is self-administered online and automatically scored, with no clinician effort. The feasibility of unsupervised cognitive screening is unclear. We adapted the Self-Administered Tasks Uncovering Risk of Neurodegeneration (SATURN) to make it suitable for self-administration and automatic scoring. 364 healthy older adults completed SATURN via a web browser, in a fully independent manner. SATURN's overall score was not modulated by gender, education, reading speed, the time of day at which the test was taken, or an individual's familiarity with technology. SATURN proved extremely portable across operating systems. Importantly, comments from participants reported satisfaction with the experience and the clarity of the instructions. SATURN represents a fast and easy screening tool that can be used for a first assessment, during a routine test or clinical evaluation, or during periodic health monitoring, in person or remotely.

Keywords: 3 to 5 keywords (or short phrases)

Cognitive screening, computer-based test, self-administered cognitive test, telecare, e-health

What the paper adds:

- a low-cost and easily accessible cognitive screening tool
- a tool in the public domain that could be adapted to specific needs

Application of study findings

- easy adaptation to different languages with consequent improvement of replicability across populations
- a proof of the feasibility of a self-administered and unsupervised screening tool
- has the potential to improve screening of cognitive decline when used as periodic health monitoring

Background

Dementia, namely the deterioration of cognitive function beyond normal biological aging, currently affects more than 55 million people worldwide, with 10 million new cases every year (World Health Organization, 2017). There is variability in how dementia manifests itself in the early stages of the disease, and patients with gradual onset may be overlooked. Screening for cognitive impairment can facilitate early detection of dementia and neurodegenerative conditions at a pre-dementia stage. Early diagnosis is extremely important for both patients and caregivers (Burns, 2012; de Vugt & Verhey, 2013), allowing them to plan for the future, initiate pharmacotherapy (Lee et al., 2004; Rasmussen & Langerman, 2019), and address problematic habits or other treatable conditions that may worsen cognitive decline (Bauer et al., 2014; Löppönen et al., 2004; Poblador-Plou et al., 2014).

Early diagnosis can be achieved using telemedicine (Koo & Vizer, 2019; Sternin et al., 2019). At least in the first phases of testing, computerized and telemedicine approaches can reduce clinician time commitment, bypassing a potential barrier to diagnosis. Telemedicine also makes it possible to screen people with physical mobility or transportation barriers to medical care, and those who do not feel the need to contact a clinician (e.g., poor insight, distorted sense of what is “age-typical”, or simply at a very early stage of cognitive decline). To facilitate early diagnosis via telemedicine delivered to an individual, there is a need for a low-cost cognitive screening, with high sensitivity at the early stages, that can be self-administered online, in an unsupervised manner, and automatically scored (García-Casal et al., 2017; Tsoy et al., 2021; Zygouris & Tsolaki, 2015). Computer testing with older adults is reliable (Sternin et al., 2019; Vaportzis et al., 2017), and computerized screening achieves comparable, or even better, diagnostic outcomes than its paper-and-pencil counterpart (Brinkman et al., 2014; Chan et al., 2021; Cyr et al., 2021). At the frontier of this screening paradigm, tele-public health initiatives might uncover cases of cognitive impairment through inexpensive and widespread testing of large cohorts of ostensibly healthy older adults – much like in-person health fairs can uncover cases of hypertension in the same population (Lucky et al., 2011). Finally, low-cost computerized screening tools could be incorporated into programs providing care for adults with limited income and resources.

Recently, the Self-Administered Tasks Uncovering Risk of Neurodegeneration (SATURN)(Bissig et al., 2020) was developed and validated for in-clinic use. This freely-available electronic

screening tool was comparable to the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) at detecting mild cognitive impairment and dementia. SATURN uses only visual stimuli (some reading capability is necessary), bypassing potential language barriers between the assessor and the patient or hearing impairment that is prevalent in the target population, and reduces some of the hardware and software requirements for remote use (e.g., no volume calibration for speakers, no multimodal stimuli to time-synchronize). SATURN is self-administered and automatically scored, two aspects that are common barriers to unsupervised use. Importantly, because SATURN is in the public domain, there were no barriers to its free adaptation to our project needs. We modified SATURN to be delivered remotely, in a completely unsupervised manner, to a relatively large cohort of healthy older adults recruited online. As usability is the weakest methodological aspect being explored when developing computerized screening tests (García-Casal et al., 2017), we also assessed SATURN's usability in this setting (Lewis, 1995, 2002).

Methods

Participants

Data were collected as part of two larger experiments (extensions of (Tagliabue et al., 2022)), in two waves (from June 2021 to September 2021 and from January 2022 to April 2022), labeled from here on sample A and sample B. Participants were recruited using Prolific (www.prolific.co). Prolific recruits individuals from, or living in, most OECD (Organisation for Economic Co-operation and Development) countries, with the exception of Turkey, Lithuania, Colombia, and Costa Rica. Prolific is also available in South Africa. The participants' pool includes individuals aged 18 and above, recruited via word of mouth (social media, flyers on university campuses, and The Prolific referral scheme (ceased March 2019), whose identity had been verified. The scheme allowed participants to invite their social network to join Prolific in return for small cash incentives for the referrer. Invitations were sent out to right-handed individuals aged between 65 to 75 years, with normal or corrected-to-normal vision, no self-declared diagnosis of mild cognitive impairment, dementia, or ongoing mental illness, and fluent in English. Participants completed the tasks on a desktop or laptop computer, and information on the operating system running on the individual devices was collected. Of the total of 408 individuals recruited, 43 were excluded for further analysis (seven because of a

mismatch between declared and recorded age, seven because of technical issues; we further restricted the sample to native English speakers only, removing 29 participants, who were otherwise considered for the larger study). Data from one participant were excluded because of an abnormal reading speed (outside three standard deviations). Thus, the final sample consisted of 364 individuals (214 female), mean age of 68.4 ± 3.1 years, range 65 to 75, of which 198 in sample A and 166 in sample B. Descriptive statistics are reported in **Table 1**. Details on the education, nationality of the sample, and CONSORT diagram are reported in the **Supplementary Materials**. The study was approved by the University of Trento Research Ethics Committee (Protocol No. 2021-041). Participants gave their informed consent at recruitment through the platform Psytoolkit (Stoet, 2010, 2017) and were compensated £2.00 GBP (~\$2.50 USD) for 20 minutes of their time.

Cognitive screening

SATURN combines scores from 20 brief tasks, cumulatively testing several cognitive domains, with a maximum score of 30 points. For tasks and scoring details, we refer the reader to the original description and testing of SATURN (Bissig et al., 2020). To make it suitable for our purposes, the original version of SATURN was replicated using PsychoPy® (v2021.2.3) (Peirce et al., 2019), translated into JavaScript, and uploaded on Pavlovia (<https://pavlovia.org/>). PsychoPy® is a Python-based free cross-platform package allowing researchers to run a wide range of behavioral experiments. PsychoPy® can run studies online using Pavlovia, a high-performance, hardware-accelerated port of the PsychoPy Python library. Both PsychoPy and Pavlovia are products of Open Science Tools Ltd. (<https://opensciencetools.org/>). While PsychoPy is free to use, Pavlovia requires a small fee (£0.24 GBP per participant or a site license) for hosting the experiment.

The following minor changes were necessary to make the tasks suitable for online unsupervised administration and scoring. First, in the original version of SATURN, participants were initially prompted to read “*close your eyes*”, and compliance was determined by the assessor to confirm that they had the necessary literacy, sensory function, and alertness to proceed with the subsequent tasks. In the current version, we first prompted individuals to choose one of two shapes on the screen (“*Click on the square to proceed*”), a test that can be automatically scored. Whereas the original version probed incidental memory by asking which phrase was read at the start, the current version later asks the participant to remember which

shape had been selected. Second, we removed the orientation question asking which state the participant was in, as it could not be scored automatically (i.e., there was no reliable way to assess the correctness of the response). We did not expect the removal of this question to impact SATURN's sensitivity, as the item showed substantial ceiling effects in scorable cognitively impaired patients in the original study (Bissig et al., 2019). Our version thus had 19 tasks and a maximum total score of 29.

Usability questionnaire

To assess the usability of the online screening tool, we administered an *ad hoc* modified version of the Post-study System Usability Questionnaire (PSSUQ) (Lewis, 1995, 2002). The PSSUQ evaluates four dimensions: overall satisfaction; system usefulness; information quality and interface quality, rated on a 7-point Likert scale (from 1 "strongly agree" to 7 "strongly disagree", also allowing for a "not applicable" option), with lower scores associated with better quality. We also administered a questionnaire developed in-house to assess the participants' familiarity with everyday technology. The two questionnaires are reported in the Supplementary Materials. SATURN data were collected as part of a larger study in two waves. In the first wave (Tagliabue et al., 2022), only individuals with a SATURN score higher than 25 would complete additional cognitive tasks over multiple sessions. In these subsequent sessions, participants also completed a series of questionnaires that included the TECH questionnaire. However, this resulted in only individuals with a SATURN score higher than 25 completing the TECH questionnaire in the first wave of data, thus biasing usability data from sample A. Therefore, in the second wave of recruitment, we decided to ask participants to complete the TECH and PSSQU questionnaires at the same time as SATURN, and to analyze only TECH and PSSQU data from sample B.

Procedure

After recruitment through the Prolific platform, participants provided their consent, confirmed their eligibility criteria, and were enrolled in the study. Individuals were then redirected to the online version of SATURN. After SATURN, only individuals in sample B filled out the PSSUQ and the technology questionnaire. Throughout the whole session, they had no contact with the experimenters and completed the screening fully independently and in an unsupervised manner. Anonymized demographic information was collected through Prolific, and data on

age, level of education, first language and gender were further confirmed through the consent form.

Statistical analysis

Statistical analysis was carried out with JASP for MAC (JASP Team, 2020). SATURN raw scores (maximum score: 29) were grouped into seven cognitive domains and summed to obtain a score for each domain (see Table 1). Participants' education was first assessed as the highest level of education completed and later recoded in actual years of school attended*. Reading speed (words per second) was calculated by dividing the reading time of instruction screens containing only text and the number of words therein. Dependency between variables was assessed with Spearman's correlation. To investigate the effect of time-of-day on performance, we coded this variable by dividing a 24h day into five intervals†. Usability and familiarity with technology were analyzed by age ranges (65 to 69 years and 70 to 75 years) and running a 2 (GENDER) x 2 (AGE RANGE) ANOVA on the questionnaires' scores. Open comments from 166 users completing the PSSQU were classified post-hoc into five categories: users who found SATURN enjoyable, interesting, or fun (FUN); users who found the instructions and tests clearly explained and easy to follow (CLEAR); users reporting on their own perceived performance (PERCEIVED PERFORMANCE); users reporting technical problems or errors (e.g., misspelled words) (TECHNICAL ISSUES); users with nothing to report (NONE). Answers for each category were counted while keeping in mind that participants could provide more than one answer. SATURN's replicability was assessed in two ways, 1) by comparing the independent sample A and B with each other via t-test and 2) via correlation analysis with data from the original SATURN validation, collected in person (Bissig et al., 2019). We decided to use correlation to assess replicability instead of a direct comparison of scores because of the differences in demographics between the samples.

* Primary education = 6 years; Secondary education (e.g. GED/GCSE) = 11 years; High school diploma/A-levels = 13 years; Technical/community college = 14 years; Undergraduate degree (BA/BSc/other) = 17 years; Graduate degree (MA/MSc/MPhil/other) = 19 years; Doctorate degree (PhD/other) = 23 years.

† 24:00 - 08:00; 08:00 - 12:00; 12:00 - 16:00; 16:00 - 20:00; 20:00 – 24:00

Additional univariate comparisons (e.g., with demographics) were run through t-tests or X^2 tests. False Discovery Rate correction for multiple comparisons was used when necessary, and two-tailed $p < 0.05$ was considered significant. SATURN source code and the anonymized data are available on the Open Science Framework: <https://osf.io/xnj5m/>.

Results

Demographic characteristics

Descriptive statistics are reported in **Table 1**, as are statistics for the total SATURN score and its sub-domains. For the entire cohort (Sample A and B), we found no gender-related differences in age, reading speed, education, total time-on-task, or overall SATURN score (t-tests, all $p_{\text{fdr}} > 0.1$). We found a significant correlation between SATURN score and age ($\rho_S(n=364) = -0.114$, $p_{\text{fdr}}=0.029$), and education ($\rho_S(n=364) = 0.163$, $p_{\text{fdr}}=0.004$). It took individuals about seven minutes to complete SATURN, with 70% of time spent completing tasks, and the remaining 30% split between breaks and reading instructions.

To further assess SATURN's robustness, we ran a comparison between two independent samples A and B, which did not differ on any demographic characteristic except for gender ($X^2_{(n=364)} = 4.95$, $p = 0.026$, with a smaller proportion of females to males in sample A than in sample B (1.2:1.9)), **Table 1**). We found no differences between the samples in SATURN scores (overall, or in each sub-domain, $p_s > 0.13$). Overall, the two independent samples A and B collected online correlated well with the sample collected in person (Bissig et al., 2019) (all significant correlations, $p < 0.001$, see **Supplementary Materials**). Qualitatively, the only difference between in-lab and online measures was on reading speed, with individuals examined in person being slower than those taking SATURN online. Descriptive statistics for each item are reported in **Table 2**. Two items (fruit words, month) were at ceiling (all subjects answered correctly).

Reading speed

Reading speed (words per second) did not significantly correlate with the overall SATURN score. We found no difference in reading speed between male and female participants. Reading speed correlated significantly with age ($\rho_S(364) = -0.170$, $p_{\text{FDR}} = 0.002$), with reading speed decreasing with increasing age, but not with education.

Time-of-day effects

As we had no control over when participants chose to complete SATURN, we investigated the potential effect of time-of-day on SATURN scores. The time of the day at which the test was taken was binned into five intervals of four hours during the day, and one interval of eight hours during the night (24:00 to 08:00). We found no association between age and the time the test was completed ($X^2_{df=4}(N = 364) = 0.774$, $p = 0.942$). We found no effect of time-of-day on SATURN scores ($F_{(4,359)} = 0.925$, $p = 0.450$, $\eta_p^2 = 0.01$), but a trend of reading speed slowing as the day progressed ($F_{(4,359)} = 2.259$, $p = 0.062$, $\eta_p^2 = 0.025$). Shifting the time ranges backward or forward by two hours did not change statistical conclusions (data not shown).

Usability

Only participants in sample B were asked to complete the TECH and PSSU questionnaires. Of those, two did not complete the questionnaires. We found a significant effect of gender on familiarity with technology ($F_{(1,160)} = 0.350$, $p = 0.022$, $\eta_p^2 = 0.03$), with women reporting to be less familiar with technological tools than men. No significant effects of age or gender were found on information quality, usefulness, interface quality, or the overall PSSUQ score. We found no correlation between the SATURN score and the PSSUQ or the TECH scores or between the PSSUQ and the TECH (all $p_s > 0.1$). Overall, participants reported high usability of SATURN (mean score on PSSUQ: TOTAL = 1.6 ± 0.8 ; Interface quality = 1.4 ± 0.8 ; usefulness = 1.8 ± 1.0 ; information quality = 1.4 ± 0.8 , for reference scores see (Lewis, 2002)). Figure 1 shows the percentage of users' comments in each category (fun (21%), clear (27%), technical issues (8%), perceived performance (22%), none (22%)). On average, comments from participants reported satisfaction with the experience and the clarity of the instructions. Few users reported some technical issues (e.g., misspelled words or, in a couple of instances, the browser being too slow to respond to their input), although these did not prevent them from

completing the tests. The list of original users' comments, grouped by category, is reported in the Supplementary Materials.

Discussion

We adapted a version of SATURN for fully unsupervised remote use through an internet browser and used it to assess a sizable cohort of healthy, English-speaking, older adults. This adds to a previous in-person validation of SATURN, which showed performance comparable to the MoCA in detecting mild cognitive impairment and dementia (Bissig et al., 2019). We assessed the feasibility of remote use through score dependency on age, education, and gender, and on the time of the day the test was taken. We further assessed the robustness of scoring across independent samples, and we compared it with a third sample previously collected in person. Finally, we quantified the overall perceived usability of the system, in three different domains: usefulness, instructions quality, and graphics quality.

There is an increasing need for screening tools to be used remotely in large cohorts. This need arises in several settings, ranging from clinical trials to public health applications.

A few other web-based or dedicated apps are available (Charalambous et al., 2020) (e.g., <https://www.aptwebstudy.org>, <https://memtrax.com/> amongst others) but they are seldom thoroughly tested, a process requiring considerable monetary and human resources. These web- or app-based tests, however, cannot be easily adapted to other uses (e.g., different populations or different environments). In addition, basic standards recommendations for medical health apps (high usability, clear language, privacy and security, and control over conflict of interests, e.g. commercialization, advertising within the app (Charalambous et al., 2020; Larson, 2018)), are not always guaranteed. SATURN, by being in the public domain, has an advantage in this sense. Beside fulfilling the requisites for basic standard quality, SATURN can be readily moved from the original clinical setting to a remote large-cohort platform, as this study demonstrates. At the time of writing, SATURN is being translated and tested in several languages. Its public domain status facilitates this development and may be essential for a reproducible screening tool that will be comparable across countries.

Despite the narrow age range covered in this study, we found that SATURN scores correlated with age and education, as expected in cognitive screening tests (Ardila et al., 2000); no gender-related differences were evident. We found that time-of-day had no impact on SATURN

scores: this is an important source of variability that we cannot control for when the test is performed remotely but could impact performance in both healthy and clinical populations (Blatter & Cajochen, 2007; Singh, Upinder et al., 2016; Wilks et al., 2021).

With the caveat that our sample had a narrow age range (65-75 years) and was comprised of self-reported healthy individuals, we showed that SATURN scores are consistent across different independent samples sharing similar demographic characteristics. Furthermore, SATURN scores correlated well with the scores of a smaller sample collected in person in the original validation study (Bissig et al., 2019). We only found a significant difference in reading speed between data collected online and in-person data (respectively 258 (sd≈78) versus 160 (sd≈44) words per minute). Differences could be due to familiarity with the interface and testing environment, the use of monetary compensation, or other unmeasured factors. Reassuringly, values were similar between samples and did not correlate with SATURN scores. Another important factor in remote unsupervised testing is represented by the dropout rate: individuals, especially those with some level of cognitive decline, might find SATURN too demanding to complete on their own. We found that only 1% of those recruited in this healthy sample dropped out before completing the tasks, a small fraction when compared to Prolific or other online recruiting platforms (Peer et al., 2017).

In the current study, individuals used only desktop or laptop computers running Linux (15/364 = 4%), Mac (42/364 = 12%) and Windows (307/364 = 84%) systems. While our sample is biased toward individuals who are familiar with computers, SATURN proved extremely portable with only 2% of participants reporting technical issues.

We found SATURN to be enjoyable, with high reported scores for usability, instruction clarity, and interface quality. These qualities are critical (García-Casal et al., 2017) for clinical testing, longitudinal monitoring, and tele-public health applications. Moreover, SATURN was not sensitive to previous experience in everyday technology (Lee Meeuw Kjoie et al., 2021).

Electronic cognitive testing in general has potential advantages over classical paper-and-pencil versions: stimuli can be easily randomized, and multiple test versions automatically implemented. The measurable variables range from accuracy to completion time, time spent on tasks, and reading time. Additionally, movement-related parameters (e.g., related to mouse use) could be recorded. All these pieces of information can be combined to better define the neuro-cognitive user's profile and, in turn, provide hints to improve early diagnosis. Tests like SATURN can be rapidly performed without supervision, and immediately and automatically

scored, which is appealing both for routine health-maintenance visits and for focused clinical evaluations.

The present adaptation of SATURN highlights a further advantage to electronic cognitive testing: testing can be delivered at home, at any time of day. More than just a convenience, this feature is necessary for tele-health clinical services, and for remote basic science applications, which blossomed during the COVID-19 pandemic. The implementation of SATURN might even be used for a "tele-public health" approach to widespread cognitive screening.

Limitations of our study include the narrow age range and the focus on self-reported healthy older adults. Further validation should include a wider age range and individuals with various levels of cognitive function to measure sensitivity and specificity. Individuals of different races/ethnicities and levels of education need to be considered in future work, as well as the use of mobile devices (smartphones and tablets). Finally, future work should focus on translating and validating SATURN in different languages, while collecting normative data specific for each population.

Following in the principles of the original SATURN implementation, we also make all SATURN materials, related to the current implementation, freely available for download ([to be added after review]), and encourage readers to use, share, and adapt SATURN without restriction.

References

- Ardila, A., Ostrosky-Solis, F., Rosselli, M., & Gómez, C. (2000). Age-Related Cognitive Decline During Normal Aging: The Complex Effect of Education. *Archives of Clinical Neuropsychology*, *15*(6), 495–513. [https://doi.org/10.1016/S0887-6177\(99\)00040-2](https://doi.org/10.1016/S0887-6177(99)00040-2)
- Bauer, K., Schwarzkopf, L., Graessel, E., & Holle, R. (2014). A claims data-based comparison of comorbidity in individuals with and without dementia. *BMC Geriatrics*, *14*(1), 10. <https://doi.org/10.1186/1471-2318-14-10>
- Bissig, D., Erten-Lyons, D., Lutsep, H., & Kaye, J. (2019). *SATURN: An inexpensive, freely-available, and fully self-administered cognitive screening test* (Version 6, p. 113728576 bytes) [Data set]. Dryad. <https://doi.org/10.5061/DRYAD.02V6WWPZR>
- Bissig, D., Kaye, J., & Erten-Lyons, D. (2020). Validation of SATURN, a free, electronic, self-administered cognitive screening test. *Alzheimer's & Dementia (New York, N. Y.)*, *6*(1), e12116. <https://doi.org/10.1002/trc2.12116>
- Blatter, K., & Cajochen, C. (2007). Circadian rhythms in cognitive performance: Methodological constraints, protocols, theoretical underpinnings. *Physiology & Behavior*, *90*(2–3), 196–208. <https://doi.org/10.1016/j.physbeh.2006.09.009>
- Brinkman, S. D., Reese, R. J., Norsworthy, L. A., Dellaria, D. K., Kinkade, J. W., Bengel, J., Brown, K., Ratka, A., & Simpkins, J. W. (2014). Validation of a Self-Administered Computerized System to Detect Cognitive Impairment in Older Adults. *Journal of Applied Gerontology*, *33*(8), 942–962. <https://doi.org/10.1177/0733464812455099>
- Burns, A. (2012). The benefits of early diagnosis of dementia. *BMJ*, *344*(may22 1), e3556–e3556. <https://doi.org/10.1136/bmj.e3556>
- Chan, J. Y. C., Yau, S. T. Y., Kwok, T. C. Y., & Tsoi, K. K. F. (2021). Diagnostic performance of digital cognitive tests for the identification of MCI and dementia: A systematic review. *Ageing Research Reviews*, *72*, 101506. <https://doi.org/10.1016/j.arr.2021.101506>
- Charalambous, A. P., Pye, A., Yeung, W. K., Leroi, I., Neil, M., Thodi, C., & Dawes, P. (2020). Tools for App- and Web-Based Self-Testing of Cognitive Impairment: Systematic Search and Evaluation. *Journal of Medical Internet Research*, *22*(1), e14551. <https://doi.org/10.2196/14551>
- Cyr, A.-A., Romero, K., & Galin-Corini, L. (2021). Web-Based Cognitive Testing of Older Adults in Person Versus at Home: Within-Subjects Comparison Study. *JMIR Aging*, *4*(1), e23384. <https://doi.org/10.2196/23384>
- de Vugt, M. E., & Verhey, F. R. J. (2013). The impact of early dementia diagnosis and intervention on informal caregivers. *Progress in Neurobiology*, *110*, 54–62. <https://doi.org/10.1016/j.pneurobio.2013.04.005>
- García-Casal, J. A., Franco-Martín, M., Perea-Bartolomé, M. V., Toribio-Guzmán, J. M., García-Moja, C., Goñi-Imizcoz, M., & Csipke, E. (2017). ELECTRONIC DEVICES FOR COGNITIVE IMPAIRMENT SCREENING: A SYSTEMATIC LITERATURE REVIEW. *International Journal of Technology Assessment in Health Care*, *33*(6), 654–673. <https://doi.org/10.1017/S0266462317000800>
- JASP Team. (2020). *JASP (Version 0.14.1)[Computer software]*. <https://jasp-stats.org/>
- Koo, B. M., & Vizer, L. M. (2019). Mobile Technology for Cognitive Assessment of Older Adults: A Scoping Review. *Innovation in Aging*, *3*(1). <https://doi.org/10.1093/geroni/igy038>
- Larson, R. S. (2018). A Path to Better-Quality mHealth Apps. *JMIR MHealth and UHealth*, *6*(7), e10414. <https://doi.org/10.2196/10414>

- Lee Meeuw Kjoie, P. R., Agelink van Rentergem, J. A., Vermeulen, I. E., & Schagen, S. B. (2021). How to Correct for Computer Experience in Online Cognitive Testing? *Assessment*, 28(5), 1247–1255. <https://doi.org/10.1177/1073191120911098>
- Lee, S., Huang, H., & Zelen, M. (2004). Early detection of disease and scheduling of screening examinations. *Statistical Methods in Medical Research*, 13(6), 443–456. <https://doi.org/10.1191/0962280204sm377ra>
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human–Computer Interaction*, 7(1), 57–78. <https://doi.org/10.1080/10447319509526110>
- Lewis, J. R. (2002). Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. *International Journal of Human–Computer Interaction*, 14(3–4), 463–488. <https://doi.org/10.1080/10447318.2002.9669130>
- Löppönen, M. K., Isoaho, R. E., Räihä, I. J., Vahlberg, T. J., Loikas, S. M., Takala, T. I., Puolijoki, H., Irjala, K. M., & Kivelä, S.-L. (2004). Undiagnosed diseases in patients with dementia— A potential target group for intervention. *Dementia and Geriatric Cognitive Disorders*, 18(3–4), 321–329. <https://doi.org/10.1159/000080126>
- Lucky, D., Turner, B., Hall, M., Lefaver, S., & de Werk, A. (2011). Blood pressure screenings through community nursing health fairs: Motivating individuals to seek health care follow-up. *Journal of Community Health Nursing*, 28(3), 119–129. <https://doi.org/10.1080/07370016.2011.588589>
- Nasreddine, Z., Phillips, N., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Poblador-Plou, B., Calderón-Larrañaga, A., Marta-Moreno, J., Hanco-Saavedra, J., Sicras-Mainar, A., Soljak, M., & Prados-Torres, A. (2014). Comorbidity of dementia: A cross-sectional study of primary care older patients. *BMC Psychiatry*, 14(1), 84. <https://doi.org/10.1186/1471-244X-14-84>
- Rasmussen, J., & Langerman, H. (2019). Alzheimer’s Disease – Why We Need Early Diagnosis. *Degenerative Neurological and Neuromuscular Disease*, 9, 123–130. <https://doi.org/10.2147/DNND.S228939>
- Singh, Upinder, Gill, M., Rice, R., Dimaano, F., Warburton, A., & Wells, M. R. (2016). Time of Day and Performance on Cognitive Tests in Patients with Mild Dementia. *Alzheimer’s & Neurodegenerative Diseases*, 2(1), 1–4. <https://doi.org/10.24966/AND-9608/100003>
- Sternin, A., Burns, A., & Owen, A. M. (2019). Thirty-Five Years of Computerized Cognitive Assessment of Aging—Where Are We Now? *Diagnostics*, 9(3), 114. <https://doi.org/10.3390/diagnostics9030114>
- Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096–1104. <https://doi.org/10.3758/BRM.42.4.1096>

- Stoet, G. (2017). PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, 44(1), 24–31. <https://doi.org/10.1177/0098628316677643>
- Tagliabue, C. F., Varesio, G., Assecondi, S., Vescovi, M., & Mazza, V. (2022). Age-related effects on online and offline learning in visuo-spatial working memory. *Aging, Neuropsychology, and Cognition*, 1–18. <https://doi.org/10.1080/13825585.2022.2054926>
- Tsoy, E., Zygouris, S., & Possin, K. (2021). Current State of Self-Administered Brief Computerized Cognitive Assessments for Detection of Cognitive Disorders in Older Adults: A Systematic Review. *The Journal of Prevention of Alzheimer's Disease*. <https://doi.org/10.14283/jpad.2021.11>
- Vaportzis, E., Giatsi Clausen, M., & Gow, A. J. (2017). Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study. *Frontiers in Psychology*, 8. <https://www.frontiersin.org/article/10.3389/fpsyg.2017.01687>
- Wilks, H., Aschenbrenner, A. J., Gordon, B. A., Balota, D. A., Fagan, A. M., Musiek, E., Balls-Berry, J., Benzinger, T. L. S., Cruchaga, C., Morris, J. C., & Hassenstab, J. (2021). Sharper in the morning: Cognitive time of day effects revealed with high-frequency smartphone testing. *Journal of Clinical and Experimental Neuropsychology*, 43(8), 825–837. <https://doi.org/10.1080/13803395.2021.2009447>
- World Health Organization. (2017). *Global action plan on the public health response to dementia 2017–2025*. World Health Organization. <https://apps.who.int/iris/handle/10665/259615>
- Zygouris, S., & Tsolaki, M. (2015). Computerized Cognitive Testing for Older Adults: A Review. *American Journal of Alzheimer's Disease & Other Dementias*®, 30(1), 13–28. <https://doi.org/10.1177/1533317514522852>

Tables

Table 1 Demographics. Mean (M), standard deviation (SD), and sample size (n) for age, years of education, gender, reading speed, total time to complete SATURN (from the “Welcome screen” to the “Goodbye screen”), and time-on-task, i.e., time spent on cognitive tasks, overall and for each sample. Statistics are also reported for the total SATURN score, and the scores associated to each sub-domain.....17

Table 2 SATURN single-item statistics. Mean (M), standard deviation (SD), and maximum score, are reported for each item. Mean and standard deviation of the time spent on each task is also reported in seconds. TMT = Trail Making Task.....18

Table 1 Demographics. Mean (M), standard deviation (SD), and sample size (n) for age, years of education, gender, reading speed, total time to complete SATURN (from the “Welcome screen” to the “Goodbye screen”), and time-on-task, i.e., time spent on cognitive tasks, overall and for each sample. Statistics are also reported for the total SATURN score, and the scores associated to each sub-domain.

	OVERALL (n=364)		SAMPLE A (n=198)		SAMPLE B (n=166)	
	M	SD	M	SD	M	SD
AGE (YEARS)	68.4	3.1	68.3	3.1	68.4	3.2
GENDER (F/M)	214 / 150	--	106 / 92	--	108 / 58	--
EDUCATION (YEARS)	15.7	3.1	15.4	3.1	16.0	3.1
READING SPEED (words/seconds)	4.3	1.3	4.3	1.3	4.2	1.2
TOTAL TIME ON TASK (minutes)	4.957	1.484	4.925	1.568	4.996	1.381
TOTAL TIME (minutes)	7.261	2.255	7.168	1.959	7.371	2.566
SATURN Total Score, out of 29	27.0	1.7	27.0	1.8	27.1	1.7
ATTENTION	6.0	0.2	6.0	0.2	6.0	0.2
INCIDENTAL	2.8	0.4	2.8	0.4	2.7	0.5
ORIENTATION	3.0	0.1	3.0	0.1	3.0	0.2
MATHS	2.8	0.7	2.7	0.8	2.9	0.5
RECALL	4.7	0.6	4.7	0.6	4.7	0.6
SPATIAL	3.6	0.7	3.6	0.7	3.5	0.7
EXECUTIVE	4.3	0.9	4.2	1.0	4.4	0.8
TECH*	--	--	--	--	19.9	3.2
PSSUQ*	--	--	--	--	1.6	0.8
* two participants did not complete the questionnaires						

Table 2 SATURN single-item statistics. Mean (M), standard deviation (SD), and maximum score, are reported for each item. Mean and standard deviation of the time spent on each task is also reported in seconds. TMT = Trail Making Task.

DOMAIN	TASK	SCORE			TIME ON TASK	
		M	SD	MAX	M	SD
ATTENTION	J-word	2.0	0.1	2.0	6.243	1.869
	Fruit words	2.0	0.0	2.0	6.027	2.482
	Number	2.0	0.2	2.0	9.646	20.863
INCIDENTAL MEMORY	Shape	0.9	0.2	1.0	7.549	3.99
	Word	1.0	0.1	1.0	7.208	2.964
	Number	0.8	0.4	1.0	8.475	5.314
ORIENTATION	Month	1.0	0.0	1.0	5.69	2.581
	Year	1.0	0.1	1.0	8.064	4.714
	Day	1.0	0.1	1.0	4.926	2.683
RECALL	Recall of five words	4.7	0.6	5.0	44.831	31.622
MATH	Sum	0.9	0.2	1.0	27.23	27.746
	Difference	1.8	0.5	2.0	10.634	8.215
VISUO- CONSTRUCTIONAL ABILITIES	Shape	1.0	0.1	1.0	9.127	5.233
	Face	0.9	0.3	1.0	21.188	14.445
	Line	0.8	0.4	1.0	19.869	13.086
	Cube	0.8	0.4	1.0	24.502	22.258
EXECUTIVE	STROOP	2.7	0.7	3.0	40.782	15.218
	TMT numbers	1.0	0.1	1.0	8.564	3.141
	TMT numbers and letters	0.6	0.5	1.0	26.882	12.33

Figures

Figure 1 Count of users' feedback in each category. Clear = number of users who found the instructions and tests clearly explained and easy to follow; Fun = number of users who found SATURN enjoyable, interesting, or fun; Perceived performance = number of users reporting on their own perceived performance; Technical issues = those reporting technical problems or errors (e.g., misspelled words in instructions or poor reactivity of the software); None = users with nothing to report. Users can give more than one answer."20

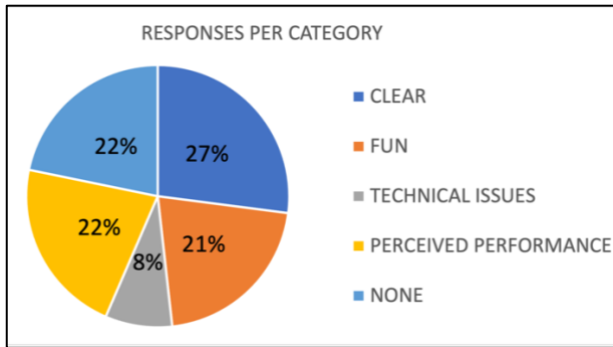


Figure 1 Percentage of users' feedback in each category. *Clear* = number of users who found the instructions and tests clearly explained and easy to follow; *Fun* = number of users who found SATURN enjoyable, interesting, or fun; *Perceived performance* = number of users reporting on their own perceived performance; *Technical issues* = those reporting technical problems or errors (e.g., misspelled words in instructions or poor reactivity of the software); *None* = users with nothing to report. Users can give more than one answer."