RESEARCH ARTICLE

Validation of SATURN, a free, electronic, self-administered cognitive screening test

David Bissig[1] | **Jeffrey Kaye[2]** | **Deniz Erten-Lyons[3]**

[1] Department of Neurology, University of California–Davis, Sacramento, California, USA

[2] Department of Neurology, Oregon Health and Science University, Portland, Oregon, USA

[3] Department of Neurology, Veterans Affairs Medical Center, Portland, Oregon, USA

Correspondence
David Bissig, Department of Neurology, University of California–Davis, 4860 Y St., Ste. 3700, Sacramento, CA 95817, USA.
Email: dpbissig@ucdavis.edu

Abstract

Background: Cognitive screening is limited by clinician time and variability in administration and scoring. We therefore developed Self-Administered Tasks Uncovering Risk of Neurodegeneration (SATURN), a free, public-domain, self-administered, and automatically scored cognitive screening test, and validated it on inexpensive (<$100) computer tablets.

Methods: SATURN is a 30-point test including orientation, word recall, and math items adapted from the Saint Louis University Mental Status test, modified versions of the Stroop and Trails tasks, and other assessments of visuospatial function and memory. English-speaking neurology clinic patients and their partners 50 to 89 years of age were given SATURN, the Montreal Cognitive Assessment (MoCA), and a brief survey about test preferences. For patients recruited from dementia clinics (n = 23), clinical status was quantified with the Clinical Dementia Rating (CDR) scale. Care partners (n = 37) were assigned CDR = 0.

Results: SATURN and MoCA scores were highly correlated ($P < .00001$; $r = 0.90$). CDR sum-of-boxes scores were well-correlated with both tests ($P < .00001$) ($r = -0.83$ and -0.86, respectively). Statistically, neither test was superior. Most participants (83%) reported that SATURN was easy to use, and most either preferred SATURN over the MoCA (47%) or had no preference (32%).

Discussion: Performance on SATURN—a fully self-administered and freely available (https://doi.org/10.5061/dryad.02v6wwpzr) cognitive screening test—is well-correlated with MoCA and CDR scores.

KEYWORDS
Alzheimer's disease, cognitive screening, computer-based test, dementia screening, psychometrics, self-administered cognitive test

1 | BACKGROUND

Dementia is prevalent and costly[1,2] but underdiagnosed: <20% of dementia cases are detected while in the "mild" stage.[3,4] Although there are dozens of cognitive screening instruments,[5] most lack some essential features when compared to successful detection tools for other conditions (like sphygmomanometry for hypertension). These include low cost, high accuracy, relative freedom from language barriers, negligible time investment from the clinician, and potential for remote use, either via telemedicine or with assistance from

non-clinicians. Legacy tools like the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA) are illustrative. These once low-cost tests have now been commercialized.[6,7] For the MoCA, commercialization via mandatory training was felt necessary to improve high variability in its administration and scoring:[6] In one study of non-clinicians, 5% of tests had administration errors and 32% had scoring errors, justifying the authors' advice to retrain testers every 2 to 3 months.[8] Furthermore, poor hearing and low vision are common in the older adults being screened.[9] Reliance on both sensory modalities complicates clinical use of the MoCA and MMSE, and an attempt to adapt the Saint Louis University Mental Status (SLUMS) test to an electronic format was limited by hearing impairment.[10] Although several legacy tests have been translated into other languages, some fluency is needed to administer the test. Finally, the 10 minutes spent administering one test can occupy an entire primary care visit, and discourages use in hospital settings where cognitive screening has prognostic value.[11–13]

Moving cognitive screening to mobile computers may mitigate the preceding issues, provided that the computer guides participants through self-administered tasks and automatically scores results. Prior efforts—reviewed elsewhere[5,14,15] and summarized in Table 1— fall short of clinical screening needs by requiring a trained test administrator,[16,17] lacking validation in dementia,[18,19] or being commercialized and restricted.[16,20–22] Responding to this gap in the literature, we developed and validated Self-Administered Tasks Uncovering Risk of Neurodegeneration (SATURN)—a free, public domain, and automatically scored cognitive screening test.

2 | METHODS

2.1 | Participants

This study was approved by the Oregon Health and Science University (OHSU) Institutional Review Board. We recruited patients and study partners from OHSU neurology clinics. Clinicians alerted study personnel when they had eligible (English-speaking adults 50 to 89 years of age) and interested patients, generating a convenience sample of 42 dyads. We prioritized recruitment from the dementia and movement disorders clinics, seeking participants with cognitive or motor impairment that might complicate use of a computer tablet. After obtaining informed consent, both dyad members participated in all study procedures, which we completed at the end of a scheduled clinic visit. We logged the primary diagnosis from each clinic visit, and when the patient was seen in our dementia clinic, he or she was assigned Clinical Dementia Rating scale global (CDR_{global}) and sum-of-boxes (CDR_{SOB}) scores. As part of a typical clinic visit, information about the study partner's activities of daily living is provided by descriptions of how he or she assists the patient. No study partners reported any significant cognitive or functional impairment. Subsequent review of partner MoCA scores supported the expectation that this group was cognitively intact.[23] They were therefore assigned CDR scores of 0.

HIGHLIGHTS

- We validate SATURN (Self-Administered Tasks Uncovering Risk of Neurodegeneration).
- SATURN is a free, self-administered, automatically scored cognitive screening test.
- SATURN is highly correlated with the Montreal Cognitive Assessment (MoCA).
- SATURN and MoCA are equally associated with Clinical Dementia Rating (CDR) score.
- All SATURN materials are freely available (https://doi.org/10.5061/dryad.02v6wwpzr).

RESEARCH IN CONTEXT

1. **Systematic review**: The authors reviewed the literature with standard sources (e.g., PubMed). There is a need for public domain, self-administered, and automatically scored cognitive screening tests, but we find no such examples that have been validated to screen for persons with dementia. Pertinent citations, including recent review articles on cognitive screening tests, are provided in the article.

2. **Interpretation**: We develop and validate a cognitive screening test that satisfies the unmet need defined by the literature. It performs favorably when compared to the Montreal Cognitive Assessment and the Clinical Dementia Rating scale.

3. **Future directions**: Because our test is in the public domain, it can be adapted or modified without restriction, which will ease deployment to clinical settings that would benefit from high-quality cognitive screening. Additional validation in diverse samples (socioeconomic, ethnic/racial, linguistic) is warranted, and draft translations to Korean, Vietnamese, and simplified Chinese are provided in a data repository.

2.2 | Study procedures

The patient and study partner were tested with SATURN and the MoCA (either version 7.1 or 7.2, which use different stimuli). For each dyad, we randomized which MoCA version was administered to the patient, and randomized the order of testing (MoCA before or after SATURN). Afterwards, participants completed a brief survey asking age, sex, years of education, and self-identified race/ethnicity, and two questions comparing the MoCA to SATURN: "Was the tablet easy to use?" ("yes" or "no"), and "Which did you prefer, the paper-and-pencil

TABLE 1 Summary of computerized self-administered[a] cognitive screening tests

| Test Name | Cognitive Domains Assessed | | | | | | | Time (min)[c] | Accommodates Hearing Loss[d] | Participants (n) Age ≥ 50 yr | Study Diagnoses | | Free to Use[e] | Reference |
	Memory	Orientation	Visuospatial	Executive	Calculation	Sustained Attention[b]	Expressive Language				MCI	Dementia		
ACAD	+					+		20	++	32			-	33
BCSD	+	+		+			+	15	++	101	+	+		34
CAMCI	+			+		+		30	+	263	+		-	35
CANS-MCI	+		+	+			+	>30		310			-	36
								50		97	+	+	-	37
CANTAB PAL	+		+						+	58	+	+	-	21
ClockMe			+	+				<2	+	20				38
										40				39
CNS Vital Signs	+			+		+		30	++	347	+	+	-	40
Cogstate Brief Battery	+					+	+	20	++	1273			-	41
										765	+	+		42
C-TOC	+		+	+	+		+	45	+	76	+	+		25
CUPDE	+	+	+		+					30				10
CUPDE2	+	+	+		+		?		?	21				43
DETECT	+			+		+		10	+	405	+	+	-	44
Dtmt			+	+					+	81	+	+		45
eCTT			+	+					++	21				43
eSAGE	+	+	+	+	+		+	18	++	66	+	+	-	20
GrayMatters	+			+				20	+	251		+	-	46
IVR	+					+	+	10		61			-	47
MicroCog/ACS	+		+	+	+	+		>30		102		+	-	48
Revere	+									153	+		-	49

(Continues)

TABLE 1 (Continued)

Test Name	Cognitive Domains Assessed							Time (min)[c]	Accommodates Hearing Loss[d]	Participants (n) Age ≥ 50 yr	Study Diagnoses		Free to Use[e]	Reference
	Memory	Orientation	Visuospatial	Executive	Calculation	Sustained Attention[b]	Expressive Language				MCI	Dementia		
SATURN	+	+	+	+	+			12	++	75	+	+	++	
——— (below, tests not yet validated in English) ———														
CADi	+	+	+	+	+	+		10		2778		+		50
CADi2	+	+	+	+	+	+		10		54		+		51
MCS	+	+	+	+	+	+	+		+	23		+		52
TPDAS	+	+	+		+			30	++	34		+		53
TPST	+	+	+	+				4		174		+		54
										>409	+	+		55

Abbreviations: ACAD, Automatic Cognitive Assessment Delivery; ACS, Assessment of Cognitive Skills; BCSD, Brief Computerized Self-screen for Dementia; CADi, Cognitive Assessment for Dementia iPad version; CADi2, Cognitive Assessment for Dementia iPad version revised; CAMCI, Computerized Assessment of Mild Cognitive Impairment; CANS-MCI, Computer-Administered Neuropsychological Screen for Mild Cognitive Impairment; CANTAB PAL, Cambridge Neuropsychological Test Automated Battery Paired Associate Learning; C-TOC, Cognitive Testing on Computer; CUPDE, Cambridge University Pen to Digital Equivalence; CUPDE2, Cambridge University Pen to Digital Equivalence revised; DETECT, Display Enhanced Testing for Cognitive impairment and Traumatic brain injury; dTMT, digital version of the Trails Making Task parts A and B; eCTT, electronic version of the Color Trails Test; eSAGE, electronic version of the Self-Administered Gerocognitive Examination; IVR, Interactive Voice Response; MCS, Mobile Cognitive Screening; SATURN, Self-Administered Tasks Uncovering Risk of Neurodegeneration; TPDAS, Touch Panel-type Dementia Assessment Scale; TPST, Touch-Panel Computer Assisted Screening Tool.

[a]We omit studies that only tested those age <50 years,[19] studies without an English-language publication (eg, the CogVal-Senior listed in a recent review[15]), and studies where a skilled operator participated in testing, as when training a participant (eg, task demonstration),[56,57] reinforcing instructions,[58] or similar.[59,60] For brevity, only one or two publications are selected for each test. The total number of participants, and inclusion (or not) of patients with dementia and mild cognitive impairment (MCI), was extracted from those specific publications. The cumulative literature experience with some tests, like MicroCog,[61] is more extensive than listed here.

[b]Includes tasks that better challenge and report on participant attention than the "Simple Attention" items in SATURN. Examples include backward digit span[51] and working memory tasks.[41,44]

[c]Where a range of times was provided, we tabled the larger value to reflect expectations for cognitively impaired participants, who are generally slower than age-matched controls. The tabled value for SATURN is above the median time for participants with a clinical dementia rating scale global score of 1.0 (Table S1). Except for Revere (for which time is difficult to report due to a 20 minutes delay built into their test procedure), values are left blank if insufficient information was provided to judge test duration.

[d]++ indicates that all testing was done without an audio device (speakers or headphones). + indicates that the study used some audio, but trivial protocol changes might make it audio-free (eg, written instructions accompanied by a recording of those instructions read aloud, for redundancy). Outside of those categories, some tests with auditory stimuli nevertheless carry many audio-free items, and truncated versions may be useful in the hearing impaired.

[e]++ indicates public domain. – indicates that the test has been commercialized, including cases using automatic speech recognition technology[47,49] and cases where we are unsure if the company is still active.[44]

test, or the tablet?" ("paper-and-pencil," "tablet," or "I felt the same about both").

2.3 | SATURN development and hardware

All code was written for VisionEgg[24] in Python 2.5.4. All stimuli were either novel, or newly adapted from public domain cognitive tests. One author (DB) produced the code and stimuli as part of his Veterans Affairs Advanced Fellowship. Therefore, SATURN is fully in the public domain. Testing was performed on low-cost (retail <$100) Ematic EWT935DK tablets running Windows 10. Stimuli appeared in the landscape orientation on the 19.7 × 11.5 cm (1024 × 600 pixel) display, with all text written in capitalized bold Arial font (each letter ≥4.5 mm tall, equivalent to ≥18 pt printed font).

Through its development, SATURN was sequentially tested in three versions. As in prior work,[10] we expected that some tasks would have limited utility when adapted for a computer tablet. We therefore included extra tasks in the first version of SATURN, with plans to remove or replace low-utility tasks in later versions. Our experience with each version is detailed in the Results section. Scoring rules were devised for each version before testing began. As in prior work,[25] each version was tested on $n > 9$ participants to gauge usability. Expecting literature-typical correlation coefficients ($r > 0.80$),[26,27] this provides adequate power ($\alpha = 0.05$; $1-\beta > 0.80$) to detect a correlation between SATURN and MoCA at each development stage. Because almost all tasks in the final version were present in the earlier versions, we also aggregated data by retroactively applying the final version's scoring rules and time limit to all three versions.

2.4 | Tasks included in SATURN's final version

Like the SLUMS and MoCA, SATURN includes high-yield[28] tests of orientation and delayed recall of intentionally encoded words. A test of incidental memory was included to ensure that any memory deficits are not merely due to poor encoding effort. To capture non-amnestic cognitive impairment, SATURN also contains brief tests of calculation, executive function, and visuospatial function. Although SATURN does not test verbal fluency, language is assessed by estimating reading speed (time spent viewing each instruction screen divided by the number of words per screen).

For all versions of SATURN, the initial tasks were the same: First, participants must read and act on the prompt "CLOSE YOUR EYES," which was written in the smallest font used throughout, and meant to verify that vision and literacy were adequate for testing. To this end, the experimenter would offer generic encouragement and re-prompt the participant multiple times as needed. Second, participants needed to eventually complete the three "simple attention" tasks described below. Even if a participant's effort and number of errors on those tasks exceeded expectations, it served as a second check that vision and literacy were adequate for testing. Per the study protocol, any participants unable to eventually complete these tasks were deemed "unscorable" and excluded from analyses (save for post hoc intent-to-screen analyses).

For all tasks, a valid input is needed to proceed (a participant cannot skip a task, and is prompted to make a selection if this is attempted) but unless otherwise noted, the input can be incorrect. The final version of SATURN is composed of these assessments, in order of their appearance (for a video of normal operation, see https://doi.org/10.5061/dryad.02v6wwpzr): Simple attention is tested when the participant must (1) select the one word (of eight) that starts with "J", (2) pick the two nouns (of four) that are fruit, and (3) copy the number 1239 with an on-screen number pad. For each of these tasks—which also acclimate the participant to SATURN, and provide materials for later testing of incidental memory—any incorrect selection leads to a prompt to change one's selection. The participant cannot proceed without selecting the correct choice. For each of these three tasks, +2 points are awarded if the correct selection is made on the first try. Participants are then asked to remember five words—presented one at a time, for two seconds each—adapted from the free recall task on the SLUMS. Participants' *incidental* memory is tested next by asking (1) which (of six) commands they read at the start of SATURN, (2) which word (of eight) they selected earlier, and (3) to re-enter the prior four-digit number. For each of these three tasks, +1 point is awarded for a correct answer. Participants' orientation for the (1) month, (2) year, (3) day of the week, and (4) state are tested, earning +1 point for a correct answer in each of these four tasks. Except for the year, which is entered with an on-screen number pad, orientation items are selected from a list of options. Next, *memory* is tested by showing a list of 100 nouns and asking the participant to select which were the five previously studied words, earning +1 point for each correct selection. *Calculation* is tested using updated versions of the SLUMS questions, where +1 point is awarded for entering the total spent on a $60 tricycle and $7 of apples, and +2 points are awarded for entering how much out of $100 remains after that purchase. Next, *visuospatial* function is tested with an adaptation of Pintner (1919)'s picture completion task[22]: Participants are shown a large drawing (eg, a circle centered on the top corner of a pentagon) and asked to select which two of six smaller drawings compose the larger one (eg, the circle, and the pentagon). Four such large drawings are shown sequentially, and for each, +1 point is earned for selecting the correct pair of smaller drawings. Next, *executive function* is probed with 12 incongruent color-word Stroop items. To reinforce task instructions, patients are prompted to input the correct answer after any error. Each error subtracts 1 point from a possible +3 points earned for perfect Stroop performance. Finally, participants earn +1 point each for error-free completion of a mini-Trails A (1 through 5) and mini-Trails B (1 through D). The maximum total score is 30.

2.5 | Statistics

Statistics were performed in R (https://www.r-project.org, v.3.6.1). We evaluated the association between MoCA and SATURN scores with linear regression. We compared both MoCA and SATURN scores to clinical status as rated by CDR_{global} (one-way analysis of variance [ANOVA])

and CDR$_{SOB}$ (linear regression), then tested whether MoCA or SAT-URN were superior predictors of CDR$_{SOB}$ (Davidson-MacKinnon *J*-test). We similarly tested for relationships between SATURN's reading speed estimate, CDR scores, and total scores on the SATURN and MoCA. Using R's pROC library, we compared receiver-operating characteristic (ROC) curves for the sensitivity and specificity of MoCA and SATURN scores for cognitive impairment. Additional univariate comparisons (as with demographics) used linear regression, *t*-tests, or Fisher exact test. Two-tailed $P < .05$ was considered significant. Task-specific findings are detailed in the Appendix.

3 | RESULTS

We refined SATURN until its third and final version. Here we present this development process, followed by analyses of aggregated data using the final version's scoring rules.

3.1 | SATURN development, Version 1

Of seven dyads (n = 14), one study partner fell outside of our pre-defined age range, and two patients were "unscorable" (see Section 3.5). Of the remaining participants (n = 11), the initial scoring plan for SATURN yielded results strongly correlated with the MoCA ($P < .00001$; $r = 0.95$). A few problems were noted. (1) Even though the six study partners appeared cognitively normal (mean ± SD MoCA scores of 27.2 ± 1.5) only one achieved better than 33% correct on an unscored spatial memory task. Due to floor effects, it was removed from subsequent versions. (2) We took note of small aesthetic choices that would improve usability, and implemented these in subsequent phases. For instance, some elements of the Stroop task instructions were mistaken for input buttons. For all subsequent versions, we therefore simplified instructions and added three unscored "warm up" trials using non-color words (eg, "BOOK" written in red) to verify task comprehension. (3) This version of SATURN shared the clock drawing task with the MoCA, but those scores were not well-correlated ($P = .052$, $r = 0.60$): nine (82%; including all six study partners) drew a normal clock on the MoCA, compared to three (27%) on the tablet. In subsequent versions of SATURN, the clock drawing was replaced by the four-item adaptation of Pintner (1919)'s picture completion task.[29] For later analyses (and presentation in Figure 1) we re-calculated Version 1 scores to approximate the final version: Lacking the picture completion task, we awarded 4 points for the clock (with 2 points for hand position), but otherwise matched the tasks and scoring rules of the final version (including a retroactively applied time limit (see Section 3.2)). Scoring changes had no effect on the relationship between total SATURN and MoCA scores ($P = .00004$; $r = 0.93$ vs the original $r = 0.95$).

3.2 | SATURN development, Version 2

Of 13 dyads, two patients were "unscorable." Of the remaining participants (n = 24), the initial scoring plan for SATURN yielded scores strongly correlated with those from the MoCA ($P < .00001$; $r = 0.84$). Before moving on to the final development, we addressed the following issues. (1) We had been piloting a full version of the Trials B task in Versions 1 and 2. For only this task, we allowed incomplete responses, which were more common in those with cognitive impairment (67%) than in their study partners (16%) ($P = .007$). We therefore opted to remove the full Trails B task from the final version of SATURN. (2) Spatial tasks borrowed from the SLUMS—drawing an "x" in a triangle and picking the largest of three shapes—showed ceiling effects, and were removed from the final version of SATURN. (3) Especially in patients with cognitive impairment, the time spent on SATURN was felt to be excessive (for Version 2, mean ± SD of 17.9 ± 8.4 min, vs 10.8 ± 4.4 in study partners; $P = .041$). We therefore built a time limit into SATURN: The program ends when one finishes whichever tasks were started within 15 minutes. Zero points are awarded for never-started tasks, and the total is still scored out of 30 points. Recalculating Version 2 scores with the final task set and time limit (as presented in Figure 1) had no effect on the correlation between SATURN and MoCA scores ($P < .00001$; $r = 0.85$ vs the original $r = 0.84$).

3.3 | SATURN, Version 3

Of 22 dyads, four study partners fell outside of our pre-defined age range. All remaining participants (n = 40) were scorable. SATURN and MoCA scores were well correlated ($P < .00001$; $r = 0.93$). CDR$_{global}$ score and CDR$_{SOB}$ were strongly associated with MoCA (respectively, $F_{[3,25]} = 38.4$, and $r = -0.89$) and SATURN scores ($F_{[3,25]} = 21.0$; $r = -0.92$)(for all, $P < .00001$).

3.4 | Scorable participants' results aggregated from all SATURN versions

Summary demographics, diagnoses, survey results, MoCA scores, and SATURN scores from the aforementioned n = 75 scorable participants are provided in Table 2.

3.4.1 | Demographics

Compared to dementia clinic patients, study partners were similar in terms of sex ($P = .4$), and years of education ($P = .4$), but were slightly younger ($P = .039$). Race and ethnicity were not analyzed, as 93% of our participants self-identified as non-Hispanic Caucasian.

3.4.2 | Survey results

Overall, 47% of participants preferred SATURN over the MoCA, and another 32% had no preference (Table 2). Compared to dementia clinic patients, study partners were more likely to favor SATURN over MoCA, and report that SATURN was easy to use ($P < .0005$).
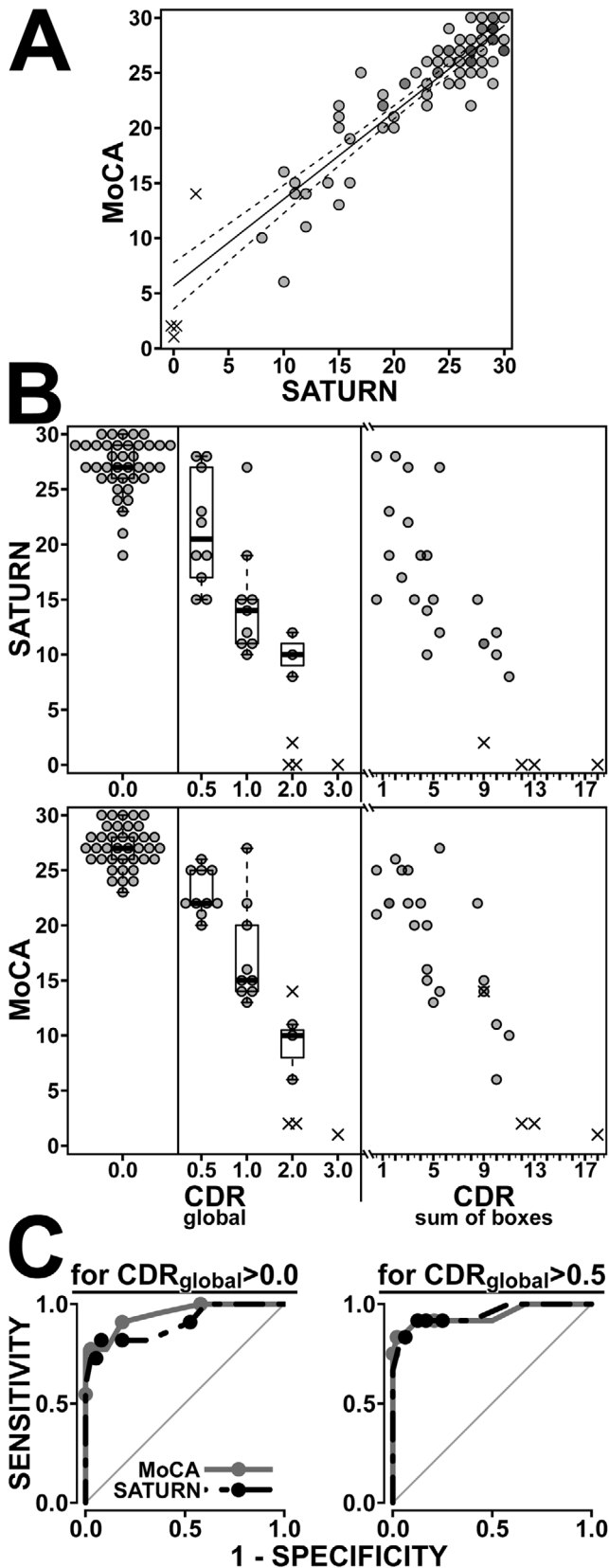
3.4.3 | Test scores in relation to demographics

MoCA scores among study partners were unrelated to age and sex (P = .2) but were lower in less-educated partners (P = .027). SATURN scores were unrelated to sex (P = .9) and years of education (P = .4), but were lower in older partners (P = .019)

3.4.4 | Test scores in relation to one-another and CDR

MoCA and SATURN scores were well correlated (P < .00001; r = 0.90). This remained true when the comparison was restricted to dementia clinic patients (n = 24 detailed in Table 2; P < .00001; r = 0.84),

correlated. Scorable participant data from all three versions of SATURN (n = 75) are shown as translucent circles, so that overlapping points appear darker. The ×s denote data from "unscorable" participants and are staggered slightly where they would otherwise overlap. (These conventions are carried through the plots in B, although, by design, the beeswarm plots for CDR_{global} have no overlapping points.) The best-fit line for the correlation among scorable participants is pictured (solid line; MoCA = 0.786 × SATURN + 5.663; r = 0.90; P < .00001) along with its 95% confidence interval (dashed lines). The 50% prediction interval (not shown) is roughly 1.5 points above and below the regression line throughout. Thus at least half of those with a SATURN score of 28 will score between a 26 and 29 on the MoCA. (B) The relationship between CDR_{global} scores and SATURN scores (top) and MoCA scores (bottom) are detailed by beeswarm plots overlaid on box-and-whisker plots. Both SATURN and MoCA were strongly associated with CDR_{global} score (analysis of variance [ANOVA]; F[3,56] = 46.8 and F[3,56] = 75.8 respectively, both P < .00001). Harmonizing this plot with Table 1, we note that those 10 participants with a CDR_{global} = 0.5 had clinical diagnoses of either mild cognitive impairment (n = 3) or mild dementia (n = 7) per American Academy of Neurology guidelines. To the right of the figure, those with nonzero CDR_{global} scores are re-plotted according to their specific CDR_{SOB} scores. Since CDR_{SOB} = 0 for all those with CDR_{global} = 0.0, their data are not re-plotted into the CDR_{SOB} plot. The relationship between CDR_{SOB} and both SATURN and MoCA was robust (linear regression; r = −0.83 and r = −0.86 respectively; P < .00001). (C) Receiver-operating characteristic (ROC) curves detail the ability of both MoCA (gray) and SATURN (black) to detect cognitive impairment. On the left, cognitive impairment is defined as CDR_{global} >0. Based on these data, it is optimal to label one cognitively impaired if one's score is <24 on SATURN (sensitivity 82%, specificity 92%) or <26 on the MoCA (sensitivity 91%, specificity 82%). Area under the curve (AUC) was similar for each test (P > .19; 0.95 for MoCA [95% CI 0.89 to 1.0], versus 0.90 for SATURN [95% CI 0.82 to 0.95]). For illustrative purposes, points are overlaid on the ROC curves at cutoffs of <22, <24, <26, and <28 for both tests. On the right, cognitive impairment is defined as CDR_{global} >0.5. Based on these data, it is optimal to label one cognitively impaired if one's score is <21 on SATURN (sensitivity 92%, specificity 88%) or <23 on the MoCA (sensitivity 92%, specificity 88%). AUC was similar for each test (P = .8; 0.94 for MoCA [95% CI 0.85 to 1.0], versus 0.95 for SATURN [95% CI 0.87 to 1.0]). Points are overlaid on the ROC curves cutoffs of <19, <21, <23, and <25 for both tests

FIGURE 1 Relationship between scores on the Montreal Cognitive Assessment (MoCA) and Self-Administered Tasks Uncovering Risk of Neurodegeneration (SATURN), and their association with clinical status, quantified by the Clinical Dementia Rating (CDR) scale. (A) SATURN and MoCA scores are strongly

TABLE 2 Demographics, mean performance, and survey results of "scorable" participants

Participant source	n	Age (years)	Sex (%women)	Education (years)	MoCA	SATURN	%Report SATURN "Easy to Use"	Test preference (%) MoCA	No Pref.	SATURN
Study partner	37	65.5 ± 9.2	59	16.3 ± 2.5	27.0 ± 1.8	27.1 ± 2.6	97	8	30	62
Dementia clinic	24	71.5 ± 10.3	46	15.5 ± 3.0	19.1 ± 6.1	17.5 ± 6.4	63	42	38	21
Not Demented[a]	4	74.3 ± 6.1	50	15.8 ± 1.6	24.8 ± 3.8	24.3 ± 4.1	75	25	75	0
Alzheimer's[b]	13	68.5 ± 11.3	46	16.4 ± 2.6	18.5 ± 5.6	16.4 ± 6.0	54	54	31	15
Other Dementia[c]	7	75.6 ± 9.2	29	13.7 ± 3.8	16.9 ± 6.6	15.7 ± 6.2	71	29	29	43
Other Clinic	14	65.9 ± 7.6	36	15.5 ± 3.6	24.8 ± 3.3	23.6 ± 3.6	79	21	29	50
Multiple Sclerosis	2	62.5 ± 7.8	100	15.0 ± 1.4	28.0 ± 1.4	26.5 ± 2.1	100	50	-	50
Essential Tremor	1	66	100	14	27	25	100	-	-	100
Parkinson	11	66.5 ± 8.2	18	15.7 ± 4.0	24.0 ± 3.3	23.0 ± 3.7	73	18	36	45

Shaded rows break down patients from each clinic into more specific diagnoses. Other than group sizes (n), data are presented as percentages or as mean ± standard deviation. The MoCA and SATURN are scored out of 30 points.

[a] Includes three patients with mild cognitive impairment, and a cognitively normal carrier of an apolipoprotein E (APOE) ε4 allele

[b] The clinical diagnosis was Alzheimer's disease for all but one, for whom cerebrospinal fluid implicated Alzheimer's disease as the cause of corticobasal syndrome

[c] Includes one case of Lewy body dementia, three cases predominantly or exclusively caused by cerebrovascular disease, and another three of mixed or uncertain cause. To simplify presentation, this row includes a patient with known dementia who was seen in general neurology clinic for peripheral neuropathy on the day of testing (and therefore not assigned CDR scores).

patients from other clinics (n = 14; $P < .00001$; $r = 0.95$; including those with Parkinson disease (n = 11; $P = .00001$; $r = 0.94$)), and study partners (n = 37; $P = .00009$; $r = 0.60$). CDR_{global} score and CDR_{SOB} were strongly associated with MoCA (respectively, $F_{[3,56]} = 75.8$, and $r = -0.86$) and SATURN ($F_{[3,56]} = 46.8$; $r = -0.83$)(all $P < .00001$). Some variance in the relationship between MoCA and CDR_{SOB} was further explained by SATURN score ($P = .048$) and vice versa ($P = .00021$) (J-tests), which indicates that SATURN and MoCA tests are complementary, without clear evidence that one is superior to the other. Multivariate analyses comparing CDR_{SOB} to either MoCA or SATURN and including demographic variables (any combination of age, education, and sex) as covariates revealed negligible impact on the relationship between CDR_{SOB} and test scores, and showed no relationship between demographics and CDR_{SOB}.

As detailed in Figure 1 and its legend, for those assigned CDR scores (n = 60), MoCA and SATURN had similar ROC curves for their ability to distinguish cognitively normal participants ($CDR_{global} = 0$) from those with any impairment ($CDR_{global} > 0$), and for their ability to distinguish those without dementia ($CDR_{global} \leq 0.5$) from those with dementia ($CDR_{global} > 0.5$).

3.4.5 | Reading speed

We compared participants' CDR scores (n = 60) to reading speed estimates (in seconds per word). Those with cognitive impairment were slower to advance through instruction screens ($r = 0.67$ for correlation with CDR_{SOB}; $F_{[3,56]} = 24.68$ for ANOVA comparison with CDR_{global}; both $P < .00001$)(see Table A.1). Multivariate analyses demonstrated that reading speed estimates complemented the 30-point test scores, explaining some additional variance in the relationship between CDR_{SOB} and SATURN ($P = .0062$) and between CDR_{SOB} and MoCA scores ($P = .014$) (J-tests). (See Appendix for additional secondary analyses of reading speed and diagnostic categories.)

3.5 | "Unscorable" participants

Across all SATURN versions, three patients were unable to read the prompt to close their eyes (all with $CDR_{global} \geq 2.0$, and MoCA ≤ 2). We initiated the program for all, since none had an ophthalmologic or educational history that would limit reading, but none earned points. The fourth unscorable patient ($CDR_{global} = 2.0$, and MoCA = 14) earned points for the first simple attention task, but erred when asked to select two fruit words from a list of four, and then seemed to confuse selection versus de-selection of responses, preventing her further advance. A post hoc intention-to-screen analyses including these four participants (age 70.8 ± 7.0 years, 15.8 ± 2.8 years education, half men, all non-Hispanic Caucasian) strengthened results: MoCA and SATURN scores remained well-correlated ($P < .00001$; $r = 0.93$). CDR_{global} score remained strongly associated with MoCA ($P < .00001$; $F_{[4,59]} = 90.1$) and SATURN ($P < .00001$; $F_{[4,59]} = 63.7$), as did CDR_{SOB} ($r = -0.92$ for MoCA, $r = -0.89$ for SATURN; both $P < .00001$) for which both

SATURN and MoCA were complementary (*J*-tests both *P* < .04). When "unscorable" patients were included in the aforementioned ROC analyses, area under the curve, sensitivity, and specificity were stable to improved.

4 | DISCUSSION

We developed and validated an electronic cognitive screening task with several desirable features. SATURN is fully self-administered and automatically scored, sparing clinician time and removing opportunities for error. Compared to prior efforts, the scope of cognitive domains tested by SATURN is favorable, and its existence in the public domain is unique (Table 1). It is strongly correlated with the previously validated MoCA, and we found no evidence that SATURN was inferior to MoCA at sorting patients by overall clinical status, as characterized with the CDR.

Given the high prevalence of hearing impairment in older adults, and based on prior work,[10] we strategically avoided auditory stimuli in SATURN. In turn, we expect SATURN will be of limited use in those with poor vision or low baseline literacy. In regular clinical practice, when a patient is unable to read the initial stimulus, an alternative and appropriately normed test should be selected. In familiar patients with good vision and high baseline literacy—as with our four "unscorable" participants—difficulty with the first stimulus may instead reflect significant cognitive impairment, and equally should trigger additional testing. The exclusive use of visually presented and automatically scored tasks may make it easier to mitigate other communication barriers: We believe that SATURN will be relatively simple to translate to other languages, and to facilitate development, we have included Korean, Vietnamese, and simplified Chinese drafts in a repository (https://doi.org/10.5061/dryad.02v6wwpzr). Language-specific versions of some tasks included in SATURN (eg, Stroop[30] and Trails[31]) have already been validated, which may further shorten development times.

One weakness of this study is the lack of demographic diversity in our sample, which was disproportionately well-educated, Caucasian, and recruited from a regional referral center. Future population-based testing in wider age ranges, wider ranges of educational achievement, and greater inclusion of racial/ethnic minorities is needed to reassure against biases. Although lacking demographic diversity, our participants were clinically diverse. We initially wondered whether patients with movement disorders or multiple sclerosis would find SATURN harder to use than the MoCA. Reassuringly, test scores remained well correlated in these patients, who tended to prefer SATURN over the MoCA. The total number of participants in this validation study is typical of prior efforts,[5,15] and it was not powered to probe for specific cognitive patterns associated with each clinical group. Like most computerized tests,[5] it is therefore not yet known if SATURN can go beyond a screening instrument, and distinguish between different types of dementia. Protocols exploring this possibility could employ additional test batteries with good discriminative power,[32] may leverage performance times recorded by SATURN (Table S1), and will benefit from greater enrollment than in the present study.

Overall, SATURN provides high-quality cognitive screening without clinician input, and is suitable for a wide variety of health care settings, including community efforts relying on non-clinicians.[8] Although the present data were collected in-person with inexpensive tablets, versions of SATURN appropriate for fully remote use—as through an internet browser—are foreseeable. Integration with electronic health records would be a worthwhile: For instance, the cognitive screening portion of a Medicare Annual Wellness Visit could be completed for most patients while sitting in the waiting room. It is doubtful that a single group or company could adapt testing to every useful setting in perpetuity. We therefore make all SATURN materials freely available for download (https://doi.org/10.5061/dryad.02v6wwpzr), and encourage readers to use, share, and adapt SATURN without restriction.

CONFLICTS OF INTEREST

Deniz Erten-Lyons, within the past year, served on the Independent Adjudication Committee for the HARMONY study by Acadia Pharmaceuticals. David Bissig and Jeffrey Kaye have no conflicts of interest.

REFERENCES

1. Wimo A, Jönsson L, Bond J, et al. The worldwide economic impact of dementia 2010. *Alzheimers Dement*. 2013;9(1):1-11.
2. https://www.alz.org/alzheimers-dementia/facts-figures Accessed April 11, 2020.
3. Boustani M, Callahan CM, Unverzagt FW, et al. Implementing a screening and diagnosis program for dementia in primary care. *J Gen Intern Med*. 2005;20(7):572-577.
4. Boise L, Neal MB, Kaye J. Dementia assessment in primary care: results from a study in three managed care systems. *J Gerontol A Biol Sci Med Sci*. 2004;59(6):M621-M626.
5. De Roeck EE, De Deyn PP, Dierckx E, Engelborghs S. Brief cognitive screening instruments for early detection of Alzheimer's disease: a systematic review. *Alzheimers Res Ther*. 2019;11(1):21.
6. https://www.mocatest.org/mandatory-moca-test-training/ Accessed April 4, 2020.
7. Kalish VB, Lerner B. Mini-mental state examination for the detection of dementia in older patients. *Am Fam Physician*. 2016;94(11):880-881.
8. Hilgeman MM, Boozer EM, Snow AL, Allen RS, Davis LL. Use of the Montreal Cognitive Assessment (MoCA) in a rural outreach program for military veterans. *J Rural Soc Sci*. 2019;34(2):2.
9. Correia C, Lopez KJ, Wroblewski KE, et al. Global sensory impairment in older adults in the United States. *J Am Geriatr Soc*. 2016;64(2):306-313.
10. Ruggeri K, Magurie A, Andrews JL, Martin E, Menon S. Are we there yet? Exploring the impact of translating cognitive tests for dementia using mobile technology in an aging population. *Front Aging Neurosci*. 2016;8:21.
11. Bissig D, DeCarli CS. Global & Community Health: Brief in-hospital cognitive screening anticipates complex admissions and may detect dementia. *Neurology*. 2019;92(13):631-634.
12. Inouye SK, Peduzzi PN, Robison JT, Hughes JS, Horwitz RI, Concato J. Importance of functional measures in predicting mortality among older hospitalized patients. *JAMA*. 1998;279(15):1187-1193.

13. Witlox J, Eurelings LS, de Jonghe JF, Kalisvaart KJ, Eikelenboom P, van Gool WA. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. *JAMA*. 2010;304(4):443-451.

14. Koo BM, Vizer LM. Mobile technology for cognitive assessment of older adults: a scoping review. *Innov Aging*. 2019;3(1):igy038.

15. García-Casal JA, Franco-Martín M, Perea-Bartolomé MV, et al. Electronic devices for cognitive impairment screening: a systematic literature review. *Int J Technol Assess Health Care*. 2017;33(6):654-673.

16. Berg JL, Durant J, Léger GC, Cummings JL, Nasreddine Z, Miller JB. Comparing the electronic and standard versions of the Montreal Cognitive Assessment in an outpatient memory disorders clinic: a validation study. *J Alzheimers Dis*. 2018;62(1):93-97.

17. 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.

18. Moore RC, Swendsen J, Depp CA. Applications for self-administered mobile cognitive assessments in clinical research: a systematic review. *Int J Methods Psychiatr Res*. 2017;26(4):e1562.

19. Nirjon S, Emi IA, Mondol MA, Salekin A, Stankovic JA. MOBI-COG: a mobile application for instant screening of dementia using the mini-cog test. In *Proceedings of the Wireless Health 2014 on National Institutes of Health*. 2014:1-7.

20. Scharre DW, Chang SI, Nagaraja HN, Vrettos NE, Bornstein RA. Digitally translated Self-Administered Gerocognitive Examination (eSAGE): relationship with its validated paper version, neuropsychological evaluations, and clinical assessments. *Alzheimers Res Ther*. 2017;9(1):44.

21. Junkkila J, Oja S, Laine M, Karrasch M. Applicability of the CANTAB-PAL computerized memory test in identifying amnestic mild cognitive impairment and Alzheimer's disease. *Dement Geriatr Cogn Disord*. 2012;34(2):83-89.

22. Saxton J, Morrow L, Eschman A, Archer G, Luther J, Zuccolotto A. Computer assessment of mild cognitive impairment. *Postgrad Med*. 2009;121(2):177-185.

23. O'Sullivan M, Brennan S, Lawlor BA, Hannigan C, Robertson IH, Pertl MM. Cognitive functioning among cognitively intact dementia caregivers compared to matched self-selected and population controls. *Aging Ment Health*. 2019;23(5):566-573.

24. Straw AD. Vision Egg: An open-source library for realtime visual stimulus generation. *Front Neuroinform*. 2008;2:4.

25. Jacova C, McGrenere J, Lee HS, et al. C-TOC (Cognitive Testing on Computer): investigating the usability and validity of a novel self-administered cognitive assessment tool in aging and early dementia. *Alzheimer Dis Assoc Disord*. 2015;29(3):213-221.

26. Stewart S, O'Riley A, Edelstein B, Gould C. A preliminary comparison of three cognitive screening instruments in long term care: The MMSE, SLUMS, and MoCA. *Clin Gerontologist*. 2012;5(1):57-75.

27. Trzepacz PT, Hochstetler H, Wang S, Walker B, Saykin AJ. Relationship between the montreal cognitive assessment and mini-mental state examination for assessment of mild cognitive impairment in older adults. *BMC Geriatr*. 2015;15:107.

28. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. *Med Care*. 2002;40(9):771-781.

29. Pintner R. A non-language group intelligence test. *J Appl Psychol*. 1919;3(3):199-214.

30. Seo EH, Lee DY, Choo IH, et al. Normative study of the stroop color and word test in an educationally diverse elderly population. *Int J Geriatr Psychiatry*. 2008;23(10):1020-1027.

31. Kim K, Jang JW, Baek MJ, Kim SY. A Comparison of three types of trail making test in the Korean elderly: higher completion rate of trail making test-black and white for mild cognitive impairment. *J Alzheimers Dis Parkinsonism*. 2016;6(239):2161-0460.

32. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a frontal assessment battery at bedside. *Neurology*. 2000;55(11):1621-1626.

33. DiRosa E, Hannigan C, Brennan S, Reilly R, Rap'can V, Robertson IH. Reliability and validity of the Automatic Cognitive Assessment Delivery (ACAD). *Front Aging Neurosci*. 2014;6:34.

34. Kluger BM, Saunders LV, Hou W, et al. A brief computerized self-screen for dementia. *J Clin Exp Neuropsychol*. 2009;31(2):234-244.

35. Tierney MC, Naglie G, Upshur R, Moineddin R, Charles J, Jaakkimainen RL. Feasibility and validity of the self-administered computerized assessment of mild cognitive impairment with older primary care patients. *Alzheimer Dis Assoc Disord*. 2014;28(4):311-319.

36. Tornatore JB, Hill E, Laboff JA, McGann ME. Self-administered screening for mild cognitive impairment: initial validation of a computerized test battery. *J Neuropsychiatry Clin Neurosci*. 2005;17(1):98-105.

37. Memória CM, Yassuda MS, Nakano EY, Forlenza OV. Contributions of the Computer-Administered Neuropsychological Screen for Mild Cognitive Impairment (CANS-MCI) for the diagnosis of MCI in Brazil. *Int Psychogeriatr*. 2014;7:1-9.

38. Kim H, Hsiao CP, Do EYL. Home-based computerized cognitive assessment tool for dementia screening. *J Ambient Intelligence Smart Environ*. 2012;4(5):429-442.

39. Zhào H, Wei W, Do EYL, Huang Y. Assessing performance on digital clock drawing test in aged patients with cerebral small vessel disease. *Front. Neurol*. 2019;10:1259.

40. Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Arch Clin Neuropsychol*. 2006;21(7):623-643.

41. Perin S, Buckley RF, Pase MP, et al. Unsupervised assessment of cognition in the Healthy Brain Project: Implications for web-based registries of individuals at risk for Alzheimer's disease. *Alzheimers Dement (N Y)*. 2020;6(1):e12043.

42. Lim YY, Ellis KA, Harrington K, et al. The AIBL Research Group. Use of the CogState Brief Battery in the assessment of Alzheimer's disease related cognitive impairment in the Australian Imaging, Biomarkers and Lifestyle (AIBL) study. *J Clin Exp Neuropsychol*. 2012;34(4):345-358.

43. Maguire Á, Martin J, Jarke H, Ruggeri K. Getting closer? Differences remain in neuropsychological assessments converted to mobile devices. *Psychol Serv*. 2018;16(2):221-226.

44. Wright DW, Nevárez H, Kilgo P, et al. A novel technology to screen for cognitive impairment in the elderly. *Am J Alzheimers Dis Other Demen*. 2011;26(6):484-491.

45. Fellows RP, Dahmen J, Cook D, Schmitter-Edgecombe M. Multi-component analysis of a digital Trail Making Test. *Clin Neuropsychol*. 2017;31(1):154-167.

46. Hicks EC. Validity of GrayMatters: A Self-Administered Computerized Assessment of Alzheimer's Disease. Digital Commons @ ACU, Electronic Theses and Dissertations. 2019. Paper 144.

47. D'Arcy S, Rapcan V, Gali A, et al. A study into the automation of cognitive assessment tasks for delivery via the telephone: lessons for developing remote monitoring applications for the elderly. *Technol Health Care*. 2013;21(4):387-396.

48. Green RC, Green J, Harrison JM, Kutner MH. Screening for cognitive impairment in older individuals. Validation study of a computer-based test. *Arch Neurol*. 1994;51(8):779-786.

49. Morrison RL, Pei H, Novak G, et al. A computerized, self-administered test of verbal episodic memory in elderly patients with mild cognitive impairment and healthy participants: A randomized, crossover, validation study. *Alzheimers Dement*. 2018;10:647-656.

50. Onoda K, Hamano T, Nabika Y, et al. Validation of a new mass screening tool for cognitive impairment: cognitive assessment for dementia, iPad version. *Clin Interv Aging*. 2013;8:353-860.

51. Onoda K, Yamaguchi S. Revision of the Cognitive Assessment for Dementia, iPad Version (CADi2). *PLoS One*. 2014;9(10):e109931.

52. Zorluoglu G, Kamasak ME, Tavacioglu L, Ozanar PO. A mobile application for cognitive screening of dementia. *Comput Methods Programs Biomed*. 2014;118(2):252-262.

53. Inoue M, Jimbo D, Taniguchi M, Urakami K. Touch panel-type dementia assessment scale: a new computer-based rating scale for Alzheimer's disease. *Psychogeriatrics*. 2011;11(1):28-33.

54. Inoue M, Jinbo D, Nakamura Y, Taniguchi M, Urakami K. Development and evaluation of a computerized test battery for Alzheimer's disease screening in community-based settings. *Am J Alzheimers Dis Other Demen*. 2009;24(2):129-135.

55. Ishiwata A, Kitamura S, Nomura T, et al. Early identification of cognitive impairment and dementia: results from four years of the community consultation center. *Arch Gerontol Geriatr*. 2014;59(2):457-461.

56. Robbins TW, James M, Owen AM, Sahakian BJ, McInnes L, Rabbitt P. Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. *Dementia*. 1994;5(5):266-281.

57. Sano M, Egelko S, Ferris S, et al. Pilot study to show the feasibility of a multicenter trial of home-based assessment of people over 75 years old. *Alzheimer Dis Assoc Disord*. 2010;24(3):256-263.

58. Levinson D, Reeves D, Watson J, Harrison M. Automated neuropsychological assessment metrics (ANAM) measures of cognitive effects of Alzheimer's disease. *Arch Clin Neuropsychol*. 2005;20(3):403-408.

59. Bayer A, Phillips M, Porter G, Leonards U, Bompas A, Tales A. Abnormal inhibition of return in mild cognitive impairment: is it specific to the presence of prodromal dementia? *J Alzheimers Dis*. 2014;40(1):177-189.

60. Dougherty JH Jr, Cannon RL, Nicholas CR, et al. The computerized self test (CST): an interactive, internet accessible cognitive screening test for dementia. *J Alzheimers Dis*. 2010;20(1):185-195.

61. Elwood RW. MicroCog: assessment of cognitive functioning. *Neuropsychol Rev*. 2001;11(2):89-100.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Bissig D, Kaye J, Erten-Lyons D. Validation of SATURN, a free, electronic, self-administered cognitive screening test. *Alzheimer's Dement*. 2020;6:e12116. https://doi.org/10.1002/trc2.12116