OXFORD
UNIVERSITY PRESS

Archives of Clinical Neuropsychology 00 (2020) 1–7

Archives
of
CLINICAL
NEUROPSYCHOLOGY

Downloaded from https://academic.oup.com/acn/advance-article/doi/10.1093/arclin/acaa056/5890330 by UPMC Mercy user on 11 November 2020

Test–retest reliability of RC21X: a web-based cognitive and neuromotor performance measurement tool

Xanthia Saganis[1], Anthony J. Goreczny[1,*], George Kondraske[2], Masha Berman[3], Ian Comnick[3], Tyler Allen[1], Paul D. Nussbaum[4]

[1]*Department of Graduate Psychology, Chatham University, Pittsburgh, PA, USA*
[2]*Electrical Engineering and Bioengineering, University of Texas at Arlington, Arlington, TX, USA*
[3]*Department of Psychology, Indiana University of Pennsylvania, Indiana, PA, USA*
[4]*RC21X, Coraopolis, PA, USA*

*Corresponding author at: Chatham University, Woodland Road, Pittsburgh, PA 15215, USA.*E-mail address:* goreczny@chatham.edu (A. J. Goreczny).

Received 18 November 2019; revised 2 July 2020; Accepted 6 July 2020

Abstract

Objective: As computerized cognitive testing becomes increasingly popular in clinical and research settings, conducting studies on efficacy and psychometric properties is essential. One such program is RC21X, a web-based brain performance measurement tool. Based on empirically supported neurocognitive and neuromotor tasks, the 12-min test consists of 15 modules measuring memory, motor coordination, processing speed, and executive functioning. Because individuals may use RC21X repeatedly to track changes in cognitive performance, establishing reliability of the program is imperative. The current study examined test–retest reliability of RC21X within a 2-week period.
Method: The sample consisted of 222 individuals: 192 (86.5%) were male, and 30 (13.5%) were female. Average age was 44.06 years ($SD = 17.76$), with ages ranging from 7 to 82 years. We computed Pearson's correlation coefficients for module and composite scores to determine reliability between performance at times 1 and 2.
Results: All correlations were statistically significant ($p < .001$). The 2-week test–retest reliability for composite score was 0.72, with subtest coefficients ranging from 0.54 on an auditory memory recognition task to 0.89 on a finger tapping task. We replicated these analyses with participants' ($n = 43$) test sessions 3 and 4; we found similar results to those from test 1 and test 2 analyses, suggesting stability of results over multiple administrations.
Conclusions: Results for RC21X were comparable to existing literature that supports moderate to high reliability of other computer-based tests. Although future research needs to investigate validity of RC21X, our findings support potential applications in research, clinical use, and personal brain performance measurement.

Keywords: Test–retest reliability; Self-administered; Neurocognitive; Neuromotor

Introduction

Neurocognitive and neuromotor performance evaluation is expanding to include web-based testing. According to Bauer et al. (2012), computerized neurocognitive tests (CNT) present several advantages over pen-and-paper assessments: (1) reduced human error; (2) increased accuracy; and (3) cost-effective, accessible testing. It is also important to distinguish tests that require administration or supervision by a professional from those available for use in home or other field settings. Web-based platforms permit self- and supervised administration.

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
doi:10.1093/arclin/acaa056

The American Academy of Clinical Neuropsychology and National Academy of Neuropsychology (2012) noted the need for researchers to establish psychometrics of web-based programs and other CNTs while adhering to the same standards as traditional pen-and-paper methods (Bauer et al., 2012). Previous studies found low to moderate test–retest reliability for composite and module scores using established CNTs, including Automated Neuropsychological Assessment Metrics (ANAM), Axon (or CogSport), Concussion Sentinel, Headminder Concussion Resolution Index, and Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT; Brett & Solomon, 2017; Nakayama et al., 2014; Nelson et al., 2016; Vincent et al., 2018).

Although clinicians traditionally have used web-based tools to compare performance before and after a neurological event, utility has expanded to include repeated administrations to monitor neurocognitive changes. One relatively new web-based instrument is RC21X, which assesses neurocognitive and neuromotor functioning. This program is unique because participants self-administer the test; thus, it has broader potential applicability over other web-based assessments that require a trained administrator. Due to novelty of the test, there is a need for research on its psychometric properties. Considering individuals may use the tool repeatedly, investigating test–retest reliability is essential. The current study examined test–retest reliability of RC21X composite and module scores.

Materials and Methods

RC21X provided de-identified data from their database for this retrospective study. The university institutional review board approved this project.

Participants

Inclusion criteria required completion of all modules with valid scores on different days within 2 weeks; there were no exclusion criteria. The website terms of agreement listed conditions necessary for participation, including ability to understand directions (e.g., sensory functioning) and access to a proper testing environment (e.g., computer with a mouse). Minors required adult consent to register. RC21X was available online. This study utilized archival data from a sample of 222 participants. Based on self-reported birthdate, ages ranged from 7 to 82 years ($M = 44.06$; $SD = 17.76$) with 192 (86.5%) males and 30 (13.5%) females.

Instrument

RC21X is a 12-min test consisting of 15 modules that measure memory, motor coordination, processing speed, and executive functioning. (For module descriptions, please contact the corresponding author or view an informational video at https://www.youtube.com/watch?v=5KkQQmGmVHM&feature=youtu.be.) Twelve modules measured lower-level performance capacities. Developers incorporated computer-based (Kondraske, 1990) and web-based (Kondraske and Stewart, 2008) tests as models. These models evolved from well-established noncomputerized methods (e.g., finger tapping, digit span; Heaton et al., 1991; Potvin et al., 1985) and concepts such as Hick's law (Hick, 1952), Fitts' Law (Fitts, 1954), and Stroop Effect (Stroop, 1935). To incorporate hierarchical diversity, the remaining modules (spaceship race game, baseball catch game, and meteor dodge game) assessed executive functioning. Although presented as games to motivate and engage users, the design included sophisticated control of task difficulty and measurement methods. Participants obtained a baseline average of initial scores, referred to as their brain performance profile (BPP). RC21X participants administered test sessions themselves, and no clinical supervision was necessary.

Measurement and scoring for RC21X is based on general systems performance theory (GSPT) and the elemental resource model (ERM) for human performance (Kondraske, 2011; Venkatachalam, 2015). The ERM, combined with GSPT, conceptualizes human functioning as different performance resources. Complex tasks (e.g., driving and completing a given work task) make quantitative demands on these resources. The demand must support the performance level of interest in the task. The goal is to measure the amount of performance resources available.

Another GSPT concept is the performance capacity envelope (PCE). In a Euclidian, multidimensional performance space, each dimension represents a specific performance resource type (e.g., speed and accuracy). Previous tests have provided unidimensional scores (e.g., separate speed and accuracy scores) or applied a form of normalization that averages measures within or across tests in a battery. Averaging involves mathematical addition, and units of speed are not the same as units of accuracy. Individual RC21X module and composite scores that ascribed to PCE perspective within GSPT recognized and preserved dimensionality with multiplication. RC21X used proprietary GSPT-based algorithms to score each module by dividing an experimentally determined estimate of the largest expected score and scaling it by 100. Each score represented a "percent maximum" value ranging to 100 (with a small fraction of cases scoring greater than 100%). To determine composite

Downloaded from https://academic.oup.com/acn/advance-article/doi/10.1093/arclin/acaa056/5890330 by UPMC Mercy user on 11 November 2020

scores, RC21X developers used a proprietary algorithm to combine 15 module scores into a single overall score that ranges on a continuum to 1,000,000. Based on the multiplicative approach, participants achieved a low score if performance on any component was low, and a high score required high performance on all constituent components. Since launched, there have been various RC21X software updates. To evaluate the system, we applied algorithms from the most recent version to raw scores to obtain the computed measures used in this sample.

Procedures

Upon initiating RC21X, participants completed introductory segments, including video introduction, overview, requirements, and instructions with examples for each module. RC21X collected demographic information, including location and session type (unsupervised or supervised). Supervision did not include trained professionals, clinicians, or RC21X staff. Tests remained self-administered, and supervisors (i.e., neurocognitively intact adults) observed to ensure adherence to instructions. Next, participants completed at least two practice sessions within 72 hr of the initial session. They then had the option of additional practice or proceeding to the first test (game) session. Practice and game sessions presented modules in the same order across administrations. RC21X used the term “game” to differentiate between practice and test sessions. Prior to each task, all modules presented video demonstrations and written instructions. After the introductory stage, participants chose time and frequency of game sessions. Therefore, participants completed a different number of sessions over various time periods. For instance, one participant may have used RC21X once, and another participant could have taken it 10 times in 1 week. Participants’ use of RC21X ranged from 2 to 92 administrations ($M = 8.54$, $SD = 11.21$). To examine test–retest reliability, we analyzed participants who took their first two game sessions within 14 days of each other and on different days.

In consideration of practice and learning effects, we eliminated data from practice sessions and utilized game sessions 1 through 4. A recent meta-analysis (Scharfen et al., 2018) determined practice effects are typically strongest between first and second administrations, with a plateau generally observed after the third administration. In our sample, participants would have reached this plateau by their first game session. For means and standard deviations for each game session, see Table 1.

Data Analyses

To determine test–retest reliability (Table 2) with participants’ composite and module scores, we calculated Pearson’s correlation coefficients for game 1 and 2 sessions taken within 2 weeks of each other. We also investigated 2-week reliability using Pearson’s correlation for supervised ($n = 76$) and unsupervised ($n = 143$) sessions, again with game 1 and 2 data and including only cases wherein both sessions had the same supervision status. We repeated reliability analyses using game sessions 3 and 4 using Pearson’s correlations because of interest in test–retest reliability over subsequent use. We conducted separate analyses to determine supervised ($n = 73$) versus unsupervised ($n = 33$) reliability using game sessions 3 and 4 (Table 2). We stratified our sample into age bands: birth—25 years ($n = 22$), 26–50 years ($n = 61$), and over 50 years ($n = 82$). To determine test–retest reliability based on age, we computed Pearson’s correlation coefficients for games 1 and 2 composite scores for each age band.

Results

All analyses described in this paragraph reflect data for the entire sample (i.e., all age bands) and had a significance of $p < .001$. We found total score reliability for RC21X was 0.72. We found moderate to high reliability for all subtests that ranged from 0.53 for auditory recognition memory to 0.90 for finger tapping. We also investigated 2-week reliability separately for supervised and unsupervised sessions. For supervised sessions, overall score reliability was high ($r = 0.74$), and module reliability coefficients in the supervised group ranged from 0.26 (visual recognition memory) to 0.91 (finger tapping). For unsupervised sessions, overall score reliability was high ($r = 0.73$), and module reliability coefficients ranged from 0.48 (visuospatial memory) to 0.90 (finger tapping).

We repeated analyses using game sessions 3 and 4. All analyses described in this paragraph reflect all G3–G4 data (including all age bands) and had a significance of $p < .001$. Composite score reliability was high ($r = 0.71$). Module score reliability coefficients ranged from 0.32 (hand-eye coordination—left) to 0.84 (finger tapping). We also conducted analyses to determine supervised versus unsupervised reliability coefficients using game sessions 3 and 4. For supervised sessions, overall score reliability was moderate ($r = 0.67$), and module reliability coefficients scores in the supervised group ranged from from 0.23 (hand-eye coordination—Left) to 0.85 (numerical memory). For unsupervised sessions, overall score reliability was high ($r = 0.85$), and module reliability coefficients ranged from 0.55 (auditory recognition memory) to 0.92 (hand–eye coordination—Left).

Downloaded from https://academic.oup.com/acn/advance-article/doi/10.1093/arclin/acaa056/5890330 by UPMC Mercy user on 11 November 2020

Table 1. Means and standard deviations for games 1–4

	Test G1 M (SD)			**Retest G2 M (SD)**		
	All cases $n = 222$	**Supervised** $n = 76$	**Unsupervised** $n = 143$	**All cases** $n = 222$	**Supervised** $n = 76$	**Unsupervised** $n = 143$
Composite	33,201.78 (53,557.398)	32,018.07 (34,698.901)	34,329.87 (61,740.809)	36,611.07 (48,521.730)	40,091.51 (45,323.866)	35,275.19 (50,541.564)
Visual recognition	70.68 (25.193)	76.11 (22.764)	67.84 (26.177)	75.83 (24.571)	83.13 (20.569)	71.78 (25.817)
Auditory recognition	52.69 (22.578)	51.45 (23.024)	53.78 (22.389)	53.63 (22.994)	53.86 (22.279)	53.48 (23.657)
Finger tapping	58.57 (16.488)	58.55 (15.326)	58.64 (17.153)	58.47 (16.484)	58.70 (14.039)	58.64 (17.479)
Speed accuracy right	48.13 (23.030)	50.37 (28.494)	47.28 (19.663)	47.06 (18.187)	49.24 (20.748)	46.29 (16.618)
Speed accuracy left	44.49 (17.740)	45.45 (20.784)	44.31 (15.993)	44.90 (17.764)	47.26 (23.250)	43.97 (13.986)
1 choice visual processing	57.28 (14.913)	59.76 (16.213)	55.87 (14.121)	57.85 (13.991)	60.67 (14.031)	56.08 (13.554)
4 choice visual processing	60.51 (13.899)	62.49 (13.763)	59.36 (14.000)	60.82 (13.811)	64.00 (12.429)	59.22 (14.310)
Visuospatial memory	53.79 (27.632)	57.18 (28.904)	52.03 (26.824)	53.32 (27.576)	57.01 (29.793)	51.40 (25.982)
Shape memory	32.96 (20.512)	35.09 (22.713)	32.01 (19.308)	35.03 (21.030)	36.17 (23.516)	34.52 (19.821)
Numerical memory	36.37 (17.128)	36.16 (19.648)	36.71 (15.779)	38.09 (18.094)	37.05 (18.769)	38.92 (17.799)
Base stealing	59.86 (15.179)	60.32 (14.875)	59.52 (15.433)	59.99 (15.751)	62.72 (13.791)	58.57 (16.698)
Baseball catch	51.24 (18.414)	57.04 (19.887)	48.53 (16.788)	51.62 (18.373)	57.82 (19.494)	48.76 (16.715)
Spaceship race	39.14 (17.647)	38.11 (19.055)	39.81 (16.980)	39.79 (17.584)	40.76 (18.430)	39.49 (17.219)
Meteor dodge	84.95 (12.684)	85.87 (12.315)	84.49 (12.978)	84.88 (12.227)	85.67 (13.310)	84.63 (11.685)
Sit stand tap	32.23 (19.702)	36.70 (21.554)	29.99 (18.203)	33.07 (21.035)	38.22 (23.178)	30.54 (19.343)
	Test G3 M (SD)			**Retest G4 M (SD)**		
	All cases $n = 107$	**Supervised** $n = 73$	**Unsupervised** $n = 33$	**All cases** $n = 107$	**Supervised** $n = 73$	**Unsupervised** $n = 33$
Composite	33,568.16 (34,017.648)	35,364.99 (35,494.485)	30,140.67 (31,126.687)	42,671.02 (43,674.360)	46,960.88 (46,365.771)	33, 721.61 (36,700.158)
Visual recognition	72.79 (23.943)	73.84 (22.183)	70.18 (27.920)	77.81 (23.446)	80.23 (21.776)	72.33 (26.629)
Auditory recognition	50.01 (22.137)	47.77 (21.936)	54.48 (22.371)	53.60 (22.701)	52.08 (23.632)	56.42 (20.676)
Finger tapping	60.90 (13.818)	61.49 (13.873)	59.52 (14.020)	61.43 (13.781)	62.12 (14.000)	60.06 (13.569)
Speed accuracy right	52.20 (20.231)	54.84 (22.177)	46.79 (14.031)	53.43 (24.291)	56.32 (27.671)	47.42 (13.065)
Speed accuracy left	48.22 (21.114)	50.18 (24.155)	44.39 (11.481)	48.78 (23.151)	51.14 (26.485)	44.00 (12.410)
1 choice visual processing	59.86 (12.901)	61.58 (12.939)	56.12 (12.386)	60.43 (13.941)	61.97 (14.126)	57.06 (13.327)
4 choice visual processing	63.77 (11.062)	65.67 (10.164)	59.70 (12.118)	63.66 (12.117)	65.34 (11.763)	60.12 (12.434)
Visuospatial memory	52.64 (26.551)	53.90 (27.519)	50.79 (24.387)	54.94 (27.470)	56.33 (27.680)	51.52 (27.469)
Shape memory	37.80 (25.176)	39.11 (26.423)	35.85 (22.162)	37.53 (22.926)	39.67 (23.976)	32.88 (20.364)
Numerical memory	37.36 (18.910)	38.16 (19.135)	35.88 (18.819)	38.59 (19.023)	39.40 (20.353)	37.45 (15.790)
Base stealing	60.86 (13.503)	61.18 (14.025)	60.27 (12.662)	61.13 (16.139)	61.73 (16.454)	59.82 (15.844)
Baseball catch	55.29 (18.805)	59.12 (18.712)	47.18 (16.661)	56.07 (17.595)	59.01 (16.555)	50.15 (18.482)
Spaceship race	42.79 (20.011)	45.16 (20.809)	37.55 (17.634)	41.48 (18.741)	43.30 (19.077)	37.27 (17.828)
Meteor dodge	88.01 (8.999)	89.66 (8.435)	84.45 (9.418)	87.17 (10.146)	88.03 (10.198)	85.79 (9.701)
Sit stand tap	39.07 (23.294)	42.90 (24.515)	29.67 (17.102)	43.38 (26.550)	47.21 (26.757)	33.97 (23.836)

Additionally, we calculated age bands for three age groups (under 26 years of age, 26–50 years of age, and older than 50 years of age) and found that reliability for composite score using G1 and G2 data was high and significant for all age groups ($r = .931$, .903, and .770 respectively, $p < .001$ for all three correlations).

Discussion

The current study analyzed test–retest reliability of RC21X over a 2-week time period. Pearson's correlation coefficients for total scores and module scores indicated moderate to high reliability that were comparable to previous literature on other CNTs (e.g., Nelson et al., 2016). Our results suggest RC21X may be reliable for clinicians, researchers, and interested individuals to use.

A major test–retest methodological issue is time between test and retest. Some studies have calculated reliability at baseline, day 45, day 50, or 2 years (Brett & Solomon, 2017; Nakayama et al., 2014; Resch et al., 2018). Another study examined reliability within 24 hr and 1, 8, 15, and 45 days post-concussion (Nelson et al., 2016). These studies show reliability across a wide range

Downloaded from https://academic.oup.com/acn/advance-article/doi/10.1093/arclin/acaa056/5890330 by UPMC Mercy user on 11 November 2020

Table 2. Pearson's correlation for total score and module scores

	G1–G2			**G3–G4**		
	All $n = 222$	**Supervised** $n = 76$	**Unsupervised** $n = 143$	**All** $n = 107$	**Supervised** $n = 73$	**Unsupervised** $n = 33$
Composite	$r = .72^{***}$	$r = .74^{***}$	$r = .73^{***}$	$r = .71^{***}$	$r = .67^{***}$	$r = .85^{***}$
Visual recognition	$r = .59^{***}$	$r = .26^{***}$	$r = .69^{***}$	$r = .51^{***}$	$r = .35^{***}$	$r = .73^{***}$
Auditory recognition	$r = .53^{***}$	$r = .41^{***}$	$r = .60^{***}$	$r = .58^{***}$	$r = .59^{***}$	$r = .55^{***}$
Finger tapping	$r = .90^{***}$	$r = .91^{***}$	$r = .90^{***}$	$r = .84^{***}$	$r = .81^{***}$	$r = .91^{***}$
Speed accuracy right	$r = .74^{***}$	$r = .70^{***}$	$r = .79^{***}$	$r = .53^{***}$	$r = .46^{***}$	$r = .90^{***}$
Speed accuracy left	$r = .67^{***}$	$r = .54^{***}$	$r = .82^{***}$	$r = .32^{***}$	$r = .23^{***}$	$r = .92^{***}$
1 choice visual processing	$r = .69^{***}$	$r = .60^{***}$	$r = .75^{***}$	$r = .57^{***}$	$r = .49^{***}$	$r = .72^{***}$
4 choice visual processing	$r = .81^{***}$	$r = .79^{***}$	$r = .82^{***}$	$r = .75^{***}$	$r = .67^{***}$	$r = .87^{***}$
Visuospatial memory	$r = .55^{***}$	$r = .64^{***}$	$r = .48^{***}$	$r = .58^{***}$	$r = .53^{***}$	$r = .74^{***}$
Shape memory	$r = .58^{***}$	$r = .64^{***}$	$r = .53^{***}$	$r = .44^{***}$	$r = .41^{***}$	$r = .56^{***}$
Numerical memory	$r = .72^{***}$	$r = .73^{***}$	$r = .71^{***}$	$r = .83^{***}$	$r = .85^{***}$	$r = .81^{***}$
Base stealing	$r = .76^{***}$	$r = .79^{***}$	$r = .76^{***}$	$r = .72^{***}$	$r = .77^{***}$	$r = .57^{***}$
Baseball catch	$r = .80^{***}$	$r = .78^{***}$	$r = .78^{***}$	$r = .65^{***}$	$r = .55^{***}$	$r = .80^{***}$
Spaceship race	$r = .64^{***}$	$r = .66^{***}$	$r = .63^{***}$	$r = .64^{***}$	$r = .58^{***}$	$r = .76^{***}$
Meteor dodge	$r = .64^{***}$	$r = .66^{***}$	$r = .63^{***}$	$r = .67^{***}$	$r = .63^{***}$	$r = .77^{***}$
Sit stand tap	$r = .72^{***}$	$r = .69^{***}$	$r = .71^{***}$	$r = .73^{***}$	$r = .68^{***}$	$r = .84^{***}$

For all correlations, $^{***}p < .001$.

of time periods and measurement frequencies. It would be valuable to investigate RC21X reliability across various time frames and frequencies as well.

We also found moderate to high reliability among individual RC21X modules. Although many subscale reliability coefficients did not meet the threshold for clinical utility (0.70; Nelson et al., 2016; Resch et al., 2018), composite score reliability for games 1 through 4 met the standard. Previous reliability studies found similar variability in subtest scores (e.g., Vincent et al., 2018). Our results may represent a low representation of module reliability because we attempted to simulate real world use by not controlling environmental factors, which most other reliability studies control. It would be valuable to investigate underlying factors contributing to differences in test–retest reliability among various tasks.

Our additional analyses addressed supervision, with moderate to high reliability in supervised and unsupervised settings. Although differences in reliability coefficients between groups appear minor, additional research is necessary to determine if an observer affects performance on RC21X. Previous research demonstrates that third-party observation may influence performance as well as present concerns in certain settings (Kuhn and Solomon, 2014; Rezaei, et al., 2017; Shindell, et al., 2014), which future research needs to evaluate. Age band analyses revealed high and significant reliability for all age groups, offering preliminary evidence that RC21X may be reliable for all ages. Future research might explore test–retest reliability among different ages given the slightly lower reliability coefficient found in the oldest age group.

A key concern with brain performance testing is unrepresentative low scores due to low effort or purposefully falsified baseline scores to create the appearance of improvement with subsequent testing (i.e., "sandbagging"; Kuhn & Solomon, 2014). An advantage of our retrospective study design is that participants did not receive explicit instructions to complete only two sessions to observe changes. Rather, they used RC21X several times at their own pace and for personal reasons. There was no incentive to perform poorly because the purpose was not to obtain disability services or return-to-play decisions; therefore sandbagging is unlikely in our sample. Considering participants completed games 1–4 after practice sessions, their task comprehension was unlikely to influence performance reliability between sessions. Reliability coefficients were similar across analyses; thus, factors like supervision and motivation did not seem to influence performance reliability on RC21X in this sample.

Our study is unique because the data has a real-world quality with participants able to complete the test outside a laboratory setting. There was a relatively heterogeneous sample, except for the predominantly male participants. Future studies need to verify these results with a female population. Importantly, we found RC21X reliability was moderate to high even with an expected variation in environmental conditions. Future research may explore test–retest reliability in a controlled laboratory setting. Germine et al. (2019) stated that environment may inflate performance because individuals perform better in distraction-free settings. Insight into how individuals test in their daily lives may help clinicians make decisions.

Clinicians, researchers, professionals, and other individuals may use RC21X in various settings, such as evaluating employee or student performance, assessing improvement through rehabilitation, and making return-to-play decisions for injured athletes. Although the current study did not assess validity of RC21X, it does provide a foundation for future studies to address reliability and validity in a variety of settings and populations. Web-based tests like RC21X are accessible to the general population who

Downloaded from https://academic.oup.com/acn/advance-article/doi/10.1093/arclin/acaa056/5890330 by UPMC Mercy user on 11 November 2020

may not have access to research participation or specialized clinical care (e.g., rural communities). Aging adults could also benefit from RC21X for screening of mild cognitive impairment and tracking cognitive changes.

Investigating RC21X psychometric properties is important because participants self-administer the test and they do not require a clinical setting or supervision. Unlike other neuropsychological tests, RC21X does not exclusively measure performance after concussions or other head injuries and has potential value for nonclinical samples. Participants did not report health history, such as cognitive impairment (e.g., traumatic brain injury), that could affect test performance. It would be valuable for future studies to compare reliability of RC21X among individuals with various health conditions. Additionally, future studies need to establish cutoff scores by comparing healthy controls with clinical populations. Establishing clinical utility may provide a wide population with access to understanding their brain performance.

RC21X is a web-based neurocognitive and neuromotor program in the ever-growing field of self-administered brain performance measurement. RC21X stems from empirically supported and standardized neuropsychological and neuromotor tests (e.g., Heaton et al., 1991). All such tests must meet psychometric standards equivalent to traditional test methods. The strong test–retest reliability of RC21X demonstrates preliminary evidence that the tool can reliably measure and monitor brain performance for researchers, clinicians, and participants.

Downloaded from https://academic.oup.com/acn/advance-article/doi/10.1093/arclin/acaa056/5890330 by UPMC Mercy user on 11 November 2020

Acknowledgements

The authors would like to acknowledge research group members for reviewing previous drafts of the manuscript. The authors would also like to acknowledge Clarence Carlos, Mark Cavicchia, and Tom Moore from RC21X for supporting this project by providing access to data.

Funding

This paper received no funding.

Conflict of Interest

The authors declare the following conflict of interests:

Xanthia Conway has no conflict of interest.

Anthony Goreczny is a Medical Advisor and equity owner with the RC21X company and has a commission-based professional services agreement but receives no paid salary.

George Kondraske has a potential research conflict of interest due to involvement with companies Human Performance Measurement, Inc. and the RC21X company. A management plan has been created to preserve objectivity in research in accordance with UTA policy.

Masha Berman has no conflict of interest.

Ian Comnick has no conflict of interest.

Tyler Allen has no conflict of interest.

Paul Nussbaum is the Chief Medical Officer and equity owner with the RC21X company but receives no paid salary.

References

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the American academy of clinical neuropsychology and the national academy of neuropsychology. *Archives of Clinical Neuropsychology*, *27*(*3*), 362–373. https://doi.org/10.1093/arclin/acs027.

Brett, B. L., & Solomon, G. S. (2017). The influence of validity criteria on immediate post-concussion assessment and cognitive testing (ImPACT) test–retest reliability among high school athletes. *Journal of Clinical and Experimental Neuropsychology*, *39*(*3*), 286–295. https://doi.org/10.1080/13803395.2016.1224322.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(*6*), 381–391 (Reprinted in Journal of Experimental Psychology: General, 121(3):262–269, 1992).

Germine, L., Reinecke, K., & Chaytor, N. S. (2019). Digital neuropsychology: Challenges and opportunities at the intersection of science and software. *The Clinical Neuropsychologist*, *33*(*2*), 271–286. https://doi.org/10.1080/13854046.2018.1535662.

Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an expanded Halstead-Reitan battery: demographic correlations, research findings, and clinical applications*. Odessa, FL: Psychological Assessment Resources.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*, 11–26.

Kondraske, G. V. (1990). A PC-based performance measurement laboratory system. *Journal of Clinical Engineering*, *15*(*6*), 467–478.

Kondraske, G. V. (2011). General systems performance theory and its application to understanding complex system performance. *Information Knowledge Systems Management*, *10*(*1–4*), 235–259. doi: 10.3233/IKS-2012-0195.

Kondraske, G. V., & Stewart, R. M. (2008). Web-based evaluation of parkinson's disease subjects: Objective performance capacity measurements and subjective characterization profiles. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 799–802. doi: 10.1109/IEMBS.2008.4649273.

Kuhn, A. W., & Solomon, G. S. (2014). Supervision and computerized neurocognitive baseline test performance in high school athletes: An initial investigation. *Journal of Athletic Training*, *49*(*6*), 800–805. https://doi.org/10.4085/1062-6050-49.3.66.

Nakayama, Y., Covassin, T., Schatz, P., Nogle, S., & Kovan, J. (2014). Examination of the test-retest reliability of a computerized neurocognitive test battery. *The American Journal of Sports Medicine*, *42*(*8*), 2000–2005. https://doi.org/10.1177/0363546514535901.

Nelson, L. D., LaRoche, A. A., Pfaller, A. Y., Lerner, E. B., Hammeke, T. A., Randolph, C. et al. (2016). Prospective, head-to-head study of three computerized neurocognitive assessment tools (CNTs): Reliability and validity for the assessment of sport-related concussion. *Journal of the International Neuropsychological Society: JINS*, *22*(*1*), 24–37. https://doi.org/10.1017/S1355617715001101.

Potvin, A. R., Tourtellotte, W. W., Potvin, J. H., Kondraske, G. V., & Syndulko, K. (1985). *Quantitative examination of neurologic functions* (, Vol. *1*, p. 247). Boca Ratan: CRC Press.

Rezaei, F., Hosseini Ramaghani, N. A., & Fazio, R. L. (2017). The effect of a third party observer and trait anxiety on neuropsychological performance: The Attentional control theory (ACT) perspective. *The Clinical Neuropsychologist*, *31*(*3*), 632–643. https://doi.org/10.1080/13854046.2016.1266031.

Resch, J. E., Schneider, M. W., & Munro Cullum, C. (2018). The test-retest reliability of three computerized neurocognitive tests used in the assessment of sport concussion. *International Journal of Psychophysiology*, *132*Pt A, 31–38. doi: 10.1016/j.ijpsycho.2017.09.011.

Scharfen, J., Jansen, K., & Holling, H. (2018). Retest effects in working memory capacity tests: A meta-analysis. *Psychonomic Bulletin & Review*, *25*(*6*), 2175–2199. doi: 10.3758/s13423-018-1461-6.

Shindell, M. R., McCaffrey, R. J., & Silk-Eglit, G. M. (2014). Three's a crowd: The impact of third-party observers on neuropsychological exams. *Minority Trial Lawyer*, *12*(*4*), 5–10.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(*6*), 643–662. doi: 10.1037/h0054651.

Venkatachalam, S. (2015). RC21X web-based brain performance capacity measurement system: Preliminary evaluations (Master's thesis). *Retrieved from ProQuest Dissertations and Theses*, *1602455*.

Vincent, A. S., Roebuck-Spencer, T. M., Fuenzalida, E., & Gilliland, K. (2018). Test-retest reliability and practice effects for the ANAM general neuropsychological screening battery. *The Clinical Neuropsychologist*, *32*(*3*), 479–494. doi: 10.1080/13854046.2017.1368716.

Downloaded from https://academic.oup.com/acn/advance-article/doi/10.1093/arclin/acaa056/5890330 by UPMC Mercy user on 11 November 2020