

Original Paper

Asynchronous Remote Assessment for Cognitive Impairment: Reliability Verification of the Neurotrack Cognitive Battery

Jennifer Rae Myers¹, DPhil; Jordan M Glenn^{1,2}, DPhil; Erica N Madero¹, MSci, MPH; John Anderson¹, BSc; Rachel Mak-McCully¹, DPhil; Michelle Gray², DPhil; Joshua L Gills², MSc; John E Harrison^{3,4,5}, DPhil

¹Neurotrack Technologies Inc, Redwood City, CA, United States

²University of Arkansas, Fayetteville, AR, United States

³Metis Cognition Ltd, Kilmington Common, United Kingdom

⁴Alzheimer Center, VU Medical Center, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

⁵Institute of Psychiatry, Psychology & Neuroscience, King's College, London, United Kingdom

Corresponding Author:

Jennifer Rae Myers, DPhil
Neurotrack Technologies Inc
399 Bradford St. #101
Redwood City, CA, 21216
United States
Phone: 1 650 549 8566
Email: jennifer@neurotrack.com

Abstract

Background: As evidenced by the further reduction in access to testing during the COVID-19 pandemic, there is an urgent, growing need for remote cognitive assessment for individuals with cognitive impairment. The Neurotrack Cognitive Battery (NCB), our response to this need, was evaluated for its temporal reliability and stability as part of ongoing validation testing.

Objective: The aim of this study is to assess the temporal reliability of the NCB tests (5 total) across a 1-week period and to determine the temporal stability of these measures across 3 consecutive administrations in a single day.

Methods: For test-retest reliability, a range of 29-66 cognitively healthy participants (ages 18-68 years) completed each cognitive assessment twice, 1 week apart. In a separate study, temporal stability was assessed using data collected from 31 different cognitively healthy participants at 3 consecutive timepoints in a single day.

Results: Correlations for the assessments were between 0.72 and 0.83, exceeding the standard acceptable threshold of 0.70 for temporal reliability. Intraclass correlations ranged from 0.60 to 0.84, indicating moderate to good temporal stability.

Conclusions: These results highlight the NCB as a brief, easy-to-administer, and reliable assessment for remote cognitive testing. Additional validation research is underway to determine the full magnitude of the clinical utility of the NCB.

(*JMIR Form Res* 2022;6(2):e34237) doi: [10.2196/34237](https://doi.org/10.2196/34237)

KEYWORDS

cognition; screening; remote testing; psychometric; challenge; validation; assessment; impairment; access; reliability; stability; testing; utility

Introduction

Background

Remote cognitive assessment, through the use of digital tools, represents an efficient means for individuals to assess their levels of function, without needing to visit an in-person clinic. These tools allow for the “benchmarking” of current levels of function, as well as the detection of change over time. Monitoring changes in performance over time enables the

detection of progressive decline that might be indicative of neurological disease. It also enables the detection of improvement from interventions targeting lifestyle changes and modifiable risk factors aimed at improving cognitive health [1-3]. Remote assessment also plays a role in areas such as screening for clinical trial participation, postmarketing surveillance of licensed therapies, and the collection of data in large prospectively ascertained cohorts. Clinical research has yielded a number of candidate measures for indexing key cognitive skills, with experts largely agreeing that, in studies

of individuals living with Alzheimer disease, assessments should measure attention, memory, and executive function [4,5].

Development of the Neurotrack Cognitive Battery

Although there are other brief and computerized cognitive assessments available, such as the Cambridge Neuropsychological Test Automated Battery (CANTAB), Savonix, and BrainCheck, the Neurotrack Cognitive Battery (NCB) offers several distinct platform features to address common concerns associated with cognitive and remote assessment. These include web camera capability, objective scoring via algorithms, and the ability to be administered without the need for a trained health care professional. Regarding assessment design, there are several recommended properties of an ideal cognitive assessment tool, which includes the need to assess the full range of relevant cognitive processes, be sensitive to aging and cognitive deficits, contain equivalent versions for repeat administration, have a reasonable testing duration, and have good reliability and validity metrics [6]. The application of recommended test selection properties led us to develop digital versions of Part B of the Trail Making Test [7] as a measure of executive function, as well as a computerized novel variant of the Digit Symbol Substitution (DSST) paradigm. The DSST enjoys the virtues of brevity and reliability, as well as being a well-known general measure of cognitive function that is sensitive to subtle changes in cognition [8]. To further index these functions, we selected the Erikson flanker task and the go/no-go test [9]. To extend the coverage of the assessment to include episodic memory, we also included a paired associate learning task in which individuals are required to pair shopping list items with the associated prices. These assessments were combined with a previously validated

associative and recognition memory visual paired comparison paradigm [10]. This task is based on research conducted by Zola et al [11] and utilizes eye tracking to determine novelty preference as an index of memory.

Measures of cognitive function must be reliable, sensitive, and valid to show meaningful change over time [12]. In the early stages of test development, we have focused on ensuring the scientific validity of the Neurotrack assessments through standard psychometric testing. Thus, the aims of our research were to (1) assess the temporal reliability of the NCB tests across a 1-week period and (2) determine the temporal stability of these measures across 3 consecutive administrations in a single day.

Methods

Participants

Temporal Reliability

Participants were workers recruited through Amazon Mechanical Turk (MTurk) and Prolific, crowdsourcing websites used for research recruitment and testing. The use of MTurk and Prolific has shown to be comparable to traditional research methods and allow for greater access to hard-to-reach and diverse populations [13,14]. Up to 150 subjectively cognitively healthy participants in the United States were recruited for each assessment separately. Individuals who successfully completed an assessment at time point 1 were granted access to retake the assessment 1 week later at time point 2. Participants were compensated up to US \$2.85 for each assessment completed. Participant characteristics for each assessment are outlined in Table 1.

Table 1. Participant characteristics for temporal reliability.

Characteristics	NCB ^a assessment and participants				
	Path points, n=66	Symbol match, n=29	Arrow match, n=46	Light reaction, n=46	Item price, n=51
Age (years), mean (SD)	32.82 (7.56)	28.97 (9.08)	33.39 (7.73)	39.50 (11.34)	36.53 (10.18)
Sex, n (%)					
Female	15 (23)	12 (41)	11 (24)	19 (41)	14 (27)
Male	51 (77)	17 (59)	35 (76)	27 (59)	37 (73)
Race, n (%)					
Nonwhite	12 (18)	8 (28)	11 (24)	8 (17)	11 (22)
White	54 (82)	21 (72)	35 (76)	38 (83)	40 (78)

^aNCB: Neurotrack Cognitive Battery.

Temporal Stability

Potential participants were recruited through a research interest listserv and by word of mouth; these individuals were selected from a different pool than the original group, with no overlap present between the two. A total of 55 individuals who were subjectively cognitively healthy expressed interest in participating in the study. Of the 55 individuals, 31 completed the study in its entirety. The mean age of the participants was 51 (SD 17.61) years. Regarding other participant characteristics,

41 of the 55 participants (75%) had a college degree, 41 (75%) were female, and 21 (38%) identified as a person of color. Participants were asked to complete the entire NCB 3 consecutive times in a single day, which took approximately 60 minutes. Participants who completed the study received a US \$50 electronic gift card.

Measures and Procedure

As previously mentioned, the following measures were selected for use based on their relative ease of administration, brevity, and tendency to represent reliable measures of cognition:

- Path points is a 2-minute assessment of executive function. This task requires participants to connect dots alternating between a number and a letter in ascending order. The primary assessment score is based on completion time.
- Symbol match is a 2-minute assessment of processing speed. Participants are instructed to determine whether 2 symbols are equal or unequal based on a legend with 9 number/symbol pairs. Participants must complete as many trials as they can in 2 minutes. Primary scores are based on accuracy and speed.
- Arrow match is a 3-minute assessment of attention. This task requires participants to indicate which direction the center arrow is pointing (left or right) among 4 distractor arrows. Primary scores are based on accuracy and speed.
- Light reaction is a 3-minute assessment of inhibition. This task requires participants to respond when they see a green light and refrain from responding when they see a red light. The primary assessment score is based on accuracy and speed.
- Item price is a 3-minute assessment of associative learning. This task requires participants to learn the prices of various produce items (eg, bananas, carrots) and identify the correct

price during the recognition trials. Primary scores are based on accuracy.

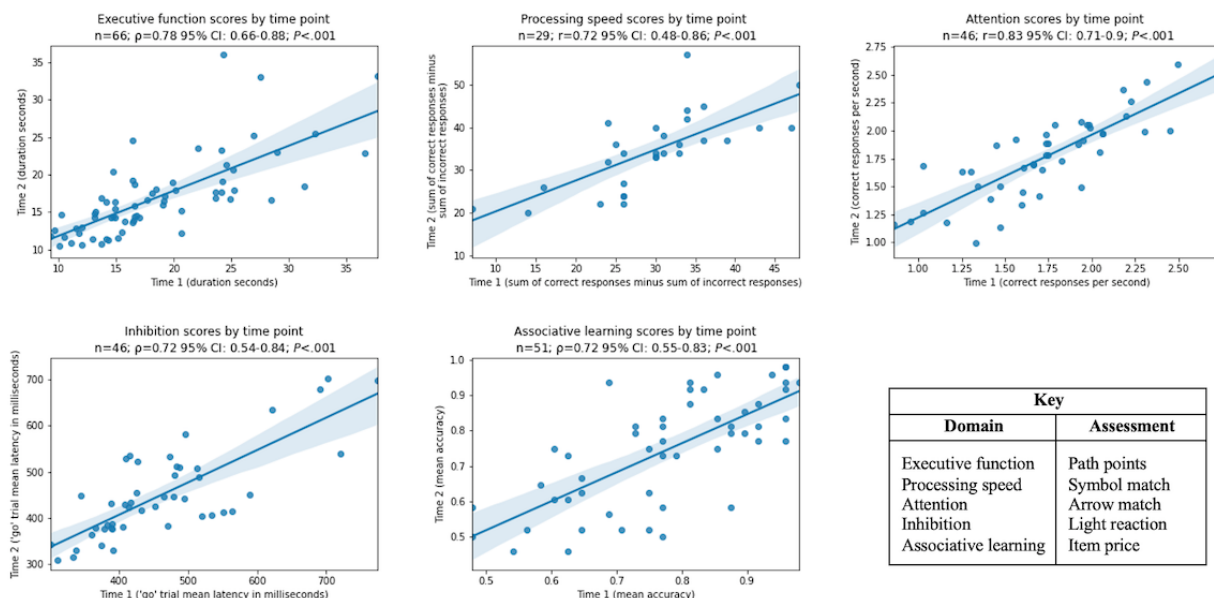
For temporal reliability and stability, data were collected separately, using new participants for each, in an unsupervised remote setting. Participants were instructed to complete the assessments when feeling rested and in a private room to optimize focus. Participants were also instructed to complete the assessments using a laptop or desktop computer with a physical keyboard, mouse, and camera, with their device charged to at least 20%, and on a stable internet connection. Study protocols were approved by the University of Arkansas Institutional Review Board and participants provided informed consent prior to study enrollment.

Results

Temporal Reliability

For each assessment, the test-retest reliability of the scores from time point 1 and 2 were assessed using both Pearson and Spearman correlations (Figure 1). In instances where a visual inspection of the data suggested a general monotonic relationship, the Spearman correlation coefficient was selected. Outliers, defined as scores more or less than 5 standard deviations from the mean, were removed from the final analyses. Correlations for the assessments ranged from 0.72 to 0.83, indicating acceptable to good test-retest reliability of the NCB.

Figure 1. Scatterplots of test-retest scores for the Neurotrack Cognitive Battery (NCB) assessments. The Pearson correlation coefficient is represented by r. The Spearman correlation coefficient represented by the Greek letter ρ .



Temporal Stability

Temporal stability was examined by calculating estimates of within-subject standard deviation (s_w) and intraclass correlation

coefficients (ICCs) for each assessment. The s_w is used to quantify measurement error in repeated measurements as a single overall measure. Results are outlined in Table 2.

Table 2. Within-subject mean and standard deviation (s_w) and Kendall rank correlation coefficient (Kendall τ) values for Neurotrack Cognitive Battery (NCB) assessments (n=31 participants).

NCB assessment	Within-subject mean	s_w	Kendall's τ	95% CI
Path points ^a	13.47	0.27	-0.42	2.48-3.57
Symbol match ^b	29.80	4.32	-0.02	3.54-5.10
Arrow match ^c	1.51	0.15	0.15	0.12-0.17
Light reaction ^d	476.69	52.25	-0.11	42.53-61.96
Item price ^e	0.74	0.08	-0.20	0.07-1.00

^aScore is measured in seconds.

^bScore is the sum of correct response minus the sum of incorrect responses.

^cScore is the number of correct responses per second.

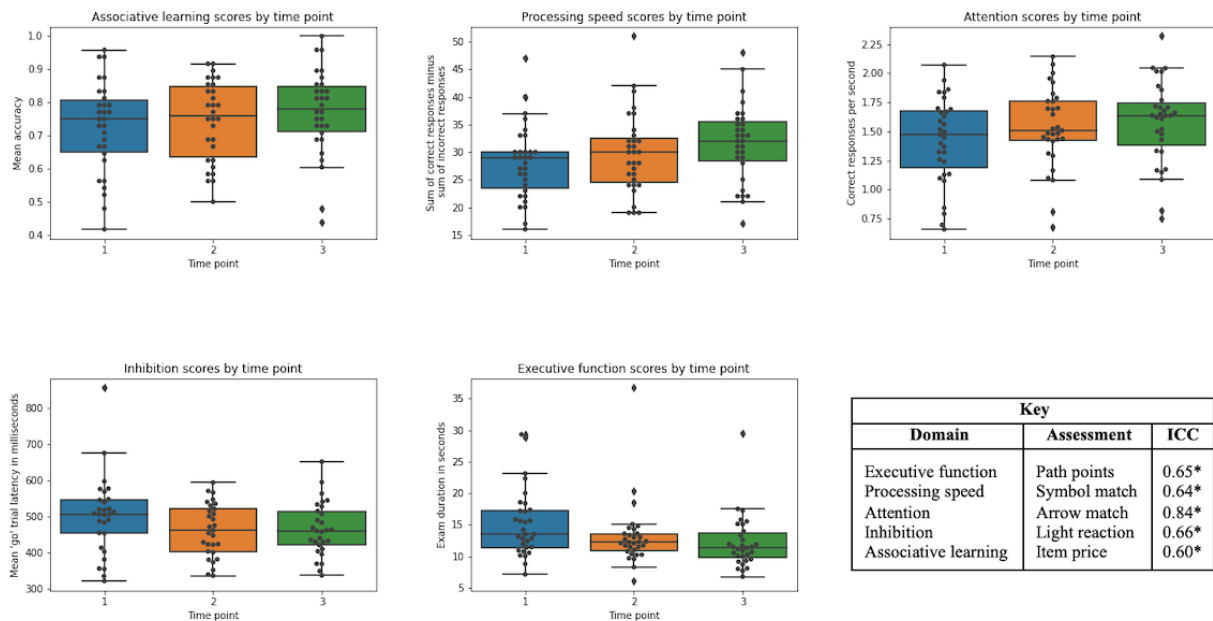
^dScore is measured in milliseconds.

^eScore is the mean accuracy of responses.

It should be noted that s_w is based on the assumption that the s_w is independent from the within-subject mean (assessed using the Kendall rank correlation coefficient [Kendall τ]). Thus, for the path points (executive function) and item price (associative learning) assessments, there is the possibility of overestimation or underestimation in how close scores are to the mean. ICCs

were also calculated to assess variation due to measurement error. The ICC (2,1) was selected as it is the recommended ICC form to use for test-retest metrics [15,16]. Correlations for the assessments ranged from 0.60 to 0.84, indicating moderate to good reliability. Boxplots for each assessment are depicted in Figure 2.

Figure 2. Boxplots representing scores from each assessment separated by time point and intraclass correlation coefficients (ICCs). The asterisk indicates $P < .001$. n=31 participants.



Discussion

Principal Findings

The results of this initial validation study indicate that the NCB is a reliable set of assessments, measuring key cognitive domains as largely accepted by the neuropsychological field. The examination of temporal reliability yielded test-retest reliability correlations which exceed the standard psychometry threshold of 0.7 for acceptable temporal reliability [17]. The NCB also

demonstrated favorable temporal stability by traditional standards given the ICC values were greater than 0.5 [15].

Limitations

The primary aim of developing the NCB has been to provide an instrument capable of reliably and remotely assessing cognition. We have sought to imbue the NCB with characteristics that facilitate the evaluation of cognitive change in group studies, as well as at the individual level. Such an approach has been advocated for some time [18] and emphasis has been placed on the need for the use of reliable measures

[12]. Although the sample was diverse, the limitations of this study are the small group sample sizes, the mean age of the participants, and the use of convenience sampling to obtain participants, which impacts the generalizability of the results. All but one assessment sample met the recommended sample size of at least 30 participants [15]; nonetheless, the use of larger sample sizes and older sample populations, as well as probability sampling methods in future validation studies is warranted.

Conclusion

This study has demonstrated that the NCB can be successfully delivered reliably and remotely, with promising clinical

implications. As evidenced by the COVID-19 pandemic, there is a critical need for feasible and valid assessments for remote testing. As it has been well established that the lack of access to cognitive assessments can result in delayed diagnoses, less effective treatment, and missed lifestyle and other health-related interventions [19], the NCB addresses this critical gap as a brief, reliable, and easy-to-administer assessment battery. Additional research regarding psychometric properties, usability, and feasibility is currently underway to determine the magnitude of the clinical validity of the NCB as part of a clinician's diagnosis process.

Acknowledgments

The authors would like to thank the participants for their time. This work was supported by funding from Neurotrack Technologies Inc.

Conflicts of Interest

JRM, JMG, ENM, JA, and RM report income and equity received from Neurotrack Technologies Inc. JEH reports personal fees from AstraZeneca PLC, AXON Neuroscience SE, Sio Gene Therapies Inc, Biogen Idec Ltd, Boehringer Ingelheim International GmbH, Signant Health, CRF Health, Eisai Co Ltd, Eli Lilly and Company, Games for Health Europe, Heptares Therapeutics Ltd, Kaasa Health GmbH, MyCognition, Neurocog, Neurotrack Technologies Inc, Novartis International AG, Nutricia, Vivoryon Therapeutics AG, Regeneron Pharmaceuticals Inc, Sanofi SA, Servier Laboratories, Takeda Pharmaceutical Company Ltd, vTv Therapeutics Inc, H. Lundbeck A/S, Compass Pathways PLC, C4X Discovery, Cognition Therapeutics Inc, AlzeCure Pharma AB, Recognify Life Sciences Inc, BlackThorn Therapeutics, Winterlight Labs, Rodin Therapeutics Inc, Lysosome Therapeutics Inc, Syndesi Therapeutics SA, Vivoryon Therapeutics NV, Neurodyn Inc, Aptinyx Inc, Athira Pharma Inc, EIP Pharma Inc, Cerecin Inc, Neurocentria Inc, CuraSen Therapeutics Inc, Biosplice Therapeutics Inc, Cognition Therapeutics Inc, ReMynd NV, Ki-Elements, and the National Health Service, outside the submitted work. MG and JLG have no conflicts of interest to disclose.

References

1. Bott N, Kumar S, Krebs C, Glenn JM, Madero EN, Juusola JL. A remote intervention to prevent or delay cognitive impairment in older adults: design, recruitment, and baseline characteristics of the Virtual Cognitive Health (VC Health) study. *JMIR Res Protoc* 2018 Aug 13;7(8):e11368 [FREE Full text] [doi: [10.2196/11368](https://doi.org/10.2196/11368)] [Medline: [30104186](https://pubmed.ncbi.nlm.nih.gov/30104186/)]
2. Glenn J, Madero EN, Gray M, Fuseya N, Ikeda M, Kawamura T, et al. Engagement with a digital platform for multimodal cognitive assessment and multidomain intervention in a Japanese population: pilot, quasi-experimental, longitudinal study. *JMIR Mhealth Uhealth* 2019 Oct 25;7(10):e15733 [FREE Full text] [doi: [10.2196/15733](https://doi.org/10.2196/15733)] [Medline: [31654567](https://pubmed.ncbi.nlm.nih.gov/31654567/)]
3. Ngandu T, Lehtisalo J, Solomon A, Levälähti E, Ahtiluoto S, Antikainen R, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *Lancet* 2015 Jun 06;385(9984):2255-2263. [doi: [10.1016/S0140-6736\(15\)60461-5](https://doi.org/10.1016/S0140-6736(15)60461-5)] [Medline: [25771249](https://pubmed.ncbi.nlm.nih.gov/25771249/)]
4. Ritchie K, Ropacki M, Albala B, Harrison J, Kaye J, Kramer J, et al. Recommended cognitive outcomes in preclinical Alzheimer's disease: consensus statement from the European Prevention of Alzheimer's Dementia project. *Alzheimers Dement* 2017 Feb;13(2):186-195. [doi: [10.1016/j.jalz.2016.07.154](https://doi.org/10.1016/j.jalz.2016.07.154)] [Medline: [27702619](https://pubmed.ncbi.nlm.nih.gov/27702619/)]
5. Vellas B, Andrieu S, Sampaio C, Coley N, Wilcock G, European Task Force Group. Endpoints for trials in Alzheimer's disease: a European task force consensus. *Lancet Neurol* 2008 May;7(5):436-450. [doi: [10.1016/S1474-4422\(08\)70087-5](https://doi.org/10.1016/S1474-4422(08)70087-5)] [Medline: [18420157](https://pubmed.ncbi.nlm.nih.gov/18420157/)]
6. Whitehouse PJ. Harmonization of Dementia Drug Guidelines (United States and Europe): a report of the International Working Group for the Harmonization for Dementia Drug Guidelines. *Alzheimer Dis Assoc Disord* 2000;14 Suppl 1:S119-S122. [doi: [10.1097/00002093-200000001-00018](https://doi.org/10.1097/00002093-200000001-00018)] [Medline: [10850740](https://pubmed.ncbi.nlm.nih.gov/10850740/)]
7. Lezak M. *Neuropsychological Assessment*, 3rd Edition. Oxford, United Kingdom: Oxford University Press; 1995.
8. Jaeger J. Digit Symbol Substitution Test: the case for sensitivity over specificity in neuropsychological testing. *J Clin Psychopharmacol* 2018 Oct;38(5):513-519 [FREE Full text] [doi: [10.1097/JCP.0000000000000941](https://doi.org/10.1097/JCP.0000000000000941)] [Medline: [30124583](https://pubmed.ncbi.nlm.nih.gov/30124583/)]
9. Donders F. On the speed of mental processes. *Acta Psychol (Amst)* 1969;30:412-431. [doi: [10.1016/0001-6918\(69\)90065-1](https://doi.org/10.1016/0001-6918(69)90065-1)] [Medline: [5811531](https://pubmed.ncbi.nlm.nih.gov/5811531/)]
10. Gills J, Glenn JM, Madero EN, Bott NT, Gray M. Validation of a digitally delivered visual paired comparison task: reliability and convergent validity with established cognitive tests. *Geroscience* 2019 Aug;41(4):441-454 [FREE Full text] [doi: [10.1007/s11357-019-00092-0](https://doi.org/10.1007/s11357-019-00092-0)] [Medline: [31463649](https://pubmed.ncbi.nlm.nih.gov/31463649/)]

11. Crutcher MD, Calhoun-Haney R, Manzanares CM, Lah JJ, Levey AI, Zola SM. Eye tracking during a visual paired comparison task as a predictor of early dementia. *Am J Alzheimers Dis Other Demen* 2009 Feb;24(3):258-266 [FREE Full text] [doi: [10.1177/1533317509332093](https://doi.org/10.1177/1533317509332093)] [Medline: [19246573](https://pubmed.ncbi.nlm.nih.gov/19246573/)]
12. Harrison JE. Measuring the mind: detecting cognitive deficits and measuring cognitive change in patients with depression. In: McIntyre RS, editor. *Cognitive Impairment in Major Depressive Disorder*. Cambridge, United Kingdom: Cambridge University Press; 2016.
13. Casler K, Bickel L, Hackett E. Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Comput Human Behav* 2013;29(6):2156-2160 [FREE Full text] [doi: [10.1016/j.chb.2013.05.009](https://doi.org/10.1016/j.chb.2013.05.009)]
14. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. *J Exp Soc Psychol* 2017 May;70:153-163. [doi: [10.1016/j.jesp.2017.01.006](https://doi.org/10.1016/j.jesp.2017.01.006)]
15. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016 Jun;15(2):155-163 [FREE Full text] [doi: [10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012)] [Medline: [27330520](https://pubmed.ncbi.nlm.nih.gov/27330520/)]
16. Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation - A discussion and demonstration of basic features. *PLoS One* 2019;14(7):e0219854 [FREE Full text] [doi: [10.1371/journal.pone.0219854](https://doi.org/10.1371/journal.pone.0219854)] [Medline: [31329615](https://pubmed.ncbi.nlm.nih.gov/31329615/)]
17. Kline P. *The Handbook of Psychological Testing*, 2nd Edition. Hove, United Kingdom: Psychology Press; 2000.
18. Harrison J, Maruff P. Measuring the mind: assessing cognitive change in clinical drug trials. *Expert Rev Clin Pharmacol* 2008 Jul;1(4):471-473. [doi: [10.1586/17512433.1.4.471](https://doi.org/10.1586/17512433.1.4.471)] [Medline: [24410548](https://pubmed.ncbi.nlm.nih.gov/24410548/)]
19. Geddes MR, O'Connell ME, Fisk JD, Gauthier S, Camicioli R, Ismail Z, Alzheimer Society of Canada Task Force on Dementia Care Best Practices for COVID - 19. Remote cognitive and behavioral assessment: report of the Alzheimer Society of Canada Task Force on dementia care best practices for COVID-19. *Alzheimers Dement (Amst)* 2020;12(1):e12111 [FREE Full text] [doi: [10.1002/dad2.12111](https://doi.org/10.1002/dad2.12111)] [Medline: [32999916](https://pubmed.ncbi.nlm.nih.gov/32999916/)]

Abbreviations

- CANTAB:** Cambridge Neuropsychological Test Automated Battery
DSST: Digit Symbol Substitution
ICC: intraclass correlation coefficient
Kendall τ : Kendall rank correlation coefficient
MTurk: Amazon Mechanical Turk
NCB: Neurotrack Cognitive Battery
 s_w : within-subject standard deviation

Edited by A Mavragani; submitted 13.10.21; peer-reviewed by M O'Connell, Y Yu; comments to author 03.11.21; revised version received 07.12.21; accepted 30.12.21; published 18.02.22

Please cite as:

Myers JR, Glenn JM, Madero EN, Anderson J, Mak-McCully R, Gray M, Gills JL, Harrison JE

Asynchronous Remote Assessment for Cognitive Impairment: Reliability Verification of the Neurotrack Cognitive Battery

JMIR Form Res 2022;6(2):e34237

URL: <https://formative.jmir.org/2022/2/e34237>

doi: [10.2196/34237](https://doi.org/10.2196/34237)

PMID:

©Jennifer Rae Myers, Jordan M Glenn, Erica N Madero, John Anderson, Rachel Mak-McCully, Michelle Gray, Joshua L Gills, John E Harrison. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 18.02.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.