

Brief Communication

Three-month Practice Effect of the National Institutes of Health Toolbox Cognition Battery in Young Healthy Adults

Leah E. Kuzmuk¹ , Alexander D. Rebchuk^{1,2} , Halina M. Deptuck³, Molly Cairncross^{4,5}, Noah D. Silverberg^{4,5} and Thalia S. Field^{1,6,7,8} 

¹Faculty of Medicine, University of British Columbia, Vancouver, Canada, ²Division of Neurosurgery, University of British Columbia, Vancouver, Canada, ³Faculty of Education, University of British Columbia, Vancouver, Canada, ⁴Department of Psychology, University of British Columbia, Vancouver, Canada, ⁵Rehabilitation Research Program, Vancouver Coastal Health Research Institute, Vancouver, Canada, ⁶Division of Neurology, University of British Columbia, Vancouver, Canada, ⁷Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, Canada and ⁸Vancouver Stroke Program, University of British Columbia, Vancouver, Canada

ABSTRACT: The National Institutes of Health Toolbox-Cognition Battery (NIHTB-CB) is a tablet-based cognitive assessment intended for individuals with neurological diseases of all ages. NIHTB-CB practice effects (PEs), however, need clarification if this measure is used to track longitudinal change. We explored the test–retest PEs on NIHTB-CB performance at 3 months in young healthy adults ($n = 22$). We examined corrected T-scores normalized for demographic factors and calculated PEs using Cohen's d . There were significant PEs for all NIHTB-CB composite scores and on 4/7 subtests. This work suggests the need to further assess NIHTB-CB PEs as this may affect the interpretation of study results incorporating this battery.

RÉSUMÉ : Effets liés à l'utilisation de la *National Institutes of Health Toolbox Cognition Battery* chez de jeunes adultes en bonne santé pendant une période de trois mois. La *National Institutes of Health Toolbox Cognition Battery* (ou NIHTB-CB) constitue une évaluation cognitive sur tablette destinée aux personnes de tous âges qui sont atteintes de maladies neurologiques. Les effets liés à l'utilisation (*practice effects*) de cet outil d'évaluation doivent cependant être clarifiés lorsqu'on s'en sert pour assurer le suivi de changements longitudinaux. Nous avons à cet égard exploré les effets des tests-retests sur les performances du NIHTB-CB au bout de trois mois chez de jeunes adultes en bonne santé ($n = 22$). Nous avons examiné les scores T corrigés et normalisés pour des facteurs démographiques en plus de calculer les effets de l'utilisation de cet outil en utilisant le d de Cohen. Ont alors émergé des effets significatifs pour tous les scores composites du NIHTB-CB ainsi que dans le cas de 4 sous-tests sur 7. Ce travail suggère donc la nécessité d'évaluer davantage les effets liés à l'utilisation de la NIHTB-CB car cela peut affecter l'interprétation des résultats des études intégrant cet outil.

Keywords: Cognition; Practice effects; Cognitive assessment; Psychometrics; Cognitive screener

(Received 28 April 2022; final revisions submitted 27 June 2022; date of acceptance 28 June 2022; First Published online 8 July 2022)

The National Institutes of Health Toolbox Cognition Battery (NIHTB-CB) is a brief tablet-based assessment of cognitive functioning designed for use across the lifespan.¹ The NIHTB-CB provides common data elements for clinical research in evaluating cognition and has been validated for use in many neurological disorders.^{2,3} The NIHTB-CB has seven subtests and produces scores normalized for age, gender, education and race–ethnicity for Fluid cognition, Crystallized cognition, and Total cognition.

Practice effects (PEs), or improvement in test performance due to repeated exposure to testing materials, have been investigated for the NIHTB-CB in a small body of literature over short- (1–5 weeks)^{4,5,6} and longer-term (15 months)⁷ intervals in middle-aged and older adults. These studies did not find consistent evidence of

PEs in these particular cohorts, over these time intervals. It is unclear if these findings can be extrapolated to younger adults, who may have a better cognitive reserve in the context of a neurological insult.⁸

Recent work from our group has demonstrated that the NIHTB-CB detects cognitive deficits in high-functioning young stroke survivors with normal scores on the Montreal Cognitive Assessment.⁹ We conducted an exploratory pilot in the healthy control group of the aforementioned study to assess possible PEs as a consideration in future research with young adults. We investigated a 3-month test–retest interval as 90-day outcomes are commonly assessed in acute stroke trials. Further, many stroke survivors will experience persisting cognitive impairment at 3

Corresponding author: Dr. Thalia Field, S169-2211 Wesbrook Mall, Vancouver BC V6T 2B5, Canada. Email: thalia.field@ubc.ca

Cite this article: Kuzmuk LE, Rebchuk AD, Deptuck HM, Cairncross M, Silverberg ND, and Field TS. (2023) Three-month Practice Effect of the National Institutes of Health Toolbox Cognition Battery in Young Healthy Adults. *The Canadian Journal of Neurological Sciences* 50: 769–772, <https://doi.org/10.1017/cjn.2022.273>

© The Author(s), 2022. Published by Cambridge University Press on behalf of Canadian Neurological Sciences Federation. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

months despite an excellent functional recovery.^{10,11} We expected higher retest performance for Fluid Cognition scores and Total Cognition. We expected stable retest performance for Crystallized Cognition scores.

We recruited healthy adults aged 18–55 years old. Further eligibility criteria included fluency in English, normal use of one's dominant hand, no history of neurological or psychiatric disease, no diagnosed learning disability, and no prior exposure to the NIHTB-CB. Participants were recruited beginning in November 2017 by advertisement at a local academic hospital. The intended sample size was a convenience sample of 50 participants, who were serving as controls for a study in young stroke survivors.⁹ This provided over 80% power a priori with a two-tailed alpha of 0.05 to detect an estimated effect size of 0.25 based on assumptions from previous test–retest work.^{6,12} Recruitment was still open in March 2020, but due to restrictions related to the COVID-19 pandemic, we completed the study early as an exploratory pilot with 22 participants.

The NIHTB-CB has seven subtests measuring five major cognitive domains: language, executive function, episodic memory, processing speed, and working memory.¹ In addition to reporting performance on individual tests, subtests are aggregated as measures of Crystallized Cognition (Picture Vocabulary and Oral Reading Recognition) and Fluid Cognition (Flanker Inhibitory Control and Attention, List Sorting Working Memory, Dimensional Change Card Sort, Pattern Comparison Processing Speed, and Picture Sequence Memory). Performance is adjusted for demographic factors, including age, education, gender, and race–ethnicity. Subtest and composite scores are reported as fully corrected T-scores (mean: 50, SD: 10).¹³

A trained research assistant administered the NIHTB-CB on a 9.7" iPad Pro (Apple, CA) in a quiet, distraction-free room. The assessment was administered in English. Administration time was approximately 30 minutes. Assessments were completed under the same test conditions with a 3-month (± 2 weeks) test–retest interval.

The study protocol was approved by the local institutional review boards and the research ethics committee at the University of British Columbia. Written informed consent was obtained by all participants.

Statistical analysis was completed using IBM SPSS Statistics version 26.0 (Armonk, NY). Demographic data are reported as descriptive statistics. Test–retest comparisons were made using paired *t*-tests and Wilcoxon rank tests (two-tailed, $p \leq 0.05$) for parametric and non-parametric data, respectively. Test–retest bivariate correlations were made using Pearson's correlation coefficient (*r*). PEs (effect sizes) were calculated using Cohen's *d* for repeated measures (within-subject version) with 95% confidence intervals and cutoffs of 0.2, 0.5, and 0.8 for small, moderate, and large effects, respectively.¹⁴ All participant data are included in the analysis.

Twenty-two participants completed both NIHTB-CB assessments over a median interval of 94 (IQR 82–106) days. Median age was 38 (IQR 34–45) years and 55% of participants were women. Self-identified race-ethnicity for most participants was either white (73%) or Asian (18%). Mean level of education was 16.0 (SD 2.5) years (Table 1).

Mean NIHTB-CB scores for Fluid, Crystallized, and Total Cognition were significantly higher at the second compared to the first administration (Table 2, see supplementary materials for participant spaghetti plots). Significant PEs were observed for all three composite cognition scores. Total Cognition had

Table 1: Participant demographics ($n = 22$)

Characteristic	
Age – median (IQR)	38 (34–45)
Female sex – no. (%)	12 (54.5)
Race – no. (%)	
White	16 (72.7)
Asian	4 (18.2)
Other	2 (9.1)
Handedness – no. (%)	
Right	19 (86.4)
Left	3 (13.6)
Education – mean (\pm SD)	15.9 (2.5)
Ethnicity/origins – no. (%)	
European	9 (40.9)
Canadian or Québécois	4 (18.2)
South Asian	1 (4.5)
East Asian	3 (13.6)
Other or mixed ethnicity	5 (22.7)
Language first learned at home – no. (%)	
English	19 (86.4)
Tagalog	2 (9.1)
Cantonese	1 (4.5)

the largest PE (Cohen's $d = 0.8$ [95% CI: 0.41–1.18], $p < 0.001$), with moderate (0.7 [0.28–1.07], $p = 0.001$) and small (0.4 [0.10–0.74], $p = 0.012$) PEs observed for Fluid and Crystallized Cognition, respectively. Moderate PEs were seen for both the Flanker Inhibitory Control (0.5 [0.14–0.86], $p = 0.007$) and Picture Sequence Memory (0.7 [0.09–1.29], $p = 0.023$) subtests, and small PEs for Picture Vocabulary (0.3 [0.04–0.52], $p = 0.032$) and Pattern Comparison Processing Speed (0.3 [0.11–0.58], $p = 0.006$) (Table 2).

In this small cohort of young, educated, healthy adults, we found a significant 3-month PEs across all composite cognition scores of the NIHTB-CB. Moderate-to-small PEs were also seen across several individual subtests, with most marked changes in Picture Sequence Memory and Flanker Inhibitory Control.

Our findings differ from previous work examining PEs in the NIHTB-CB as both Fluid and Crystallized composite scores improved on re-testing. Previous work investigating a short test–retest interval of 3 weeks found a small to moderate effect size for Fluid Cognition ($d = 0.42$) and Total Cognition ($d = 0.29$), but no significant PE for Crystallized Cognition.⁴ Our findings are unexpected. While fluid cognition represents dynamic cognitive processes, including working memory and executive function, which are more susceptible to aging or brain injury, crystallized cognition reflects cognitive processes relying on language and comprehension and tends to be stable across the lifespan with greater resilience to brain changes.^{3,4} Thus, it is possible that, despite their emphasis on testing crystallized cognition, the design of the Picture Vocabulary subtest may still be sensitive to learning effects, at least amongst our cohort and over this time interval.

In contrast to 3-week repeated administration of the NIHTB-CB, an extended test–retest period of 15 months in older adults

Table 2: NIHTB-CB Performance (fully-corrected T-scores) at test vs. retest over 3-month interval ($n = 22$)

	Test mean (SD)	Re-test mean (SD)	Pearson's r (test-retest correlation)	p -value (test-retest differences*)	Cohen's d (95% CI)
Crystallized Cognition	60 (5.9)	62.6 (6.6)	0.75	0.012	0.4 (0.10–0.74)
Picture Vocabulary [^]	59.2 (8.5)	61.7 (9.1)	0.85	0.032	0.3 (0.04–0.52)
Oral Reading Recognition	58.9 (5.5)	61.9 (6.7)	0.29	0.074	0.5 (–0.06 to 1.02)
Fluid Cognition[^]	54.7 (10.9)	61.9 (10.3)	0.66	0.001	0.7 (0.28–1.07)
Flanker Inhibitory Control and Attention	45.9 (10.3)	51.2 (10.9)	0.69	0.007	0.5 (0.14–0.86)
List Sorting Working Memory [^]	53.5 (8.9)	52.8 (9.4)	0.66	0.677	0.2 (–0.43 to 0.28)
Dimensional Change Card Sort Test [^]	55.0 (13.2)	60.3 (13.3)	0.56	0.102	0.4 (–0.03 to 0.81)
Pattern Comparison Processing Speed	54.1 (17.3)	59.9 (14.4)	0.86	0.006	0.3 (0.11–0.58)
Picture Sequence Memory [^]	57.7 (13.8)	66.2 (10.4)	0.23	0.023	0.7 (0.09–1.29)
Total Cognition	58.8 (7.2)	64.6 (7.4)	0.70	<0.001	0.8 (0.41–1.18)

* p -value for t -test and Wilcoxon rank test.

[^]Nonparametric data which used Wilcoxon rank test for test–retest comparisons, all other comparisons were made using a paired t -test.

showed no significant PEs on any of the composite cognition scores, and a small effect size for the Dimensional Change Card Sort task.⁷ These differences are not unexpected as PEs are typically more prominent with a shorter test-retest interval, and with higher frequency testing.¹⁵ Our results are in line with PEs reported in older adults over a 3–5-week retest interval where significant PEs were seen for both Fluid and Total Cognition scores, as well as the Picture Sequence Memory task.⁵ However, it is challenging to compare PE trends between our younger adult sample and older adults as fluid cognition is known to decrease across adulthood in parallel with age-related neurobiological changes.¹⁶ Improved characterization of PEs helps to minimize overestimation of cognitive recovery, or, conversely, underestimation of cognitive deterioration, over time. Our findings suggest that three-month PEs, even on Crystallized Cognition performance, may be a potential consideration in interpreting longitudinal changes in NIHTB-CB scores.

There are limitations to our study. We had a smaller-than-anticipated sample size due to early termination with the COVID-19 pandemic. Importantly, the demographics of our study group, who are primarily white with post-secondary education, are not representative of the general population. However, cultural and educational backgrounds are thought to impact crystallized cognitive ability more than fluid cognition, which would not explain the moderate PEs for the Fluid Cognition score reported here.⁴ Additionally, early-life education, as well as education in mid-to-late adulthood has been shown to improve crystallized cognition, but does not influence working memory, executive function, or other fluid cognitive abilities.¹⁷ The homogenous nature of our sample is further demonstrated by the fact that some of the standard deviations for the NIHTB-CB T-scores were small, for Total Cognition, and with a leptokurtic distribution. This may contribute to the moderate-to-large observed PEs on composite scores, despite nonsignificant effect sizes on many individual subtests. We found similar distributions in a previous study examining 3-week PEs in in-person versus virtual test conditions in healthy controls who had similarly high levels of education.⁶ We also acknowledge that, although the 3-month interval was chosen in considering usual timepoints for trials in acute stroke, different follow-up time intervals will be preferred for studies of different neurological and psychiatric conditions. Finally, we did not account for other factors such as sleep and stress, which may confound performance.

Despite the limitations of this pilot study, however, given that the NIHTB-CB is increasingly used as an outcome measure in studies of neurological and psychiatric disease, we feel it is important to draw attention to potential considerations of PEs if the NIHTB-CB is to be used as a repeated measure. Future confirmatory work with larger and more diverse cohorts is warranted.

Our findings suggest that the NIHTB-CB may have lower test-retest reliability in short-term repeated administrations in young, educated, healthy adults. Although preliminary, this work suggests that PEs may need to be considered in studies repeating the NIHTB-CB at 3-month intervals, and that even Fluid Cognition subtests may be subject to PEs in some circumstances. This work may be informative for clinical trials or observational studies using the NIHTB-CB to assess longitudinal cognitive outcomes.

Supplementary Material. For supplementary material accompanying this paper visit <https://doi.org/10.1017/cjn.2022.273>

Acknowledgments. Zoe O'Neill, BSc., medical student at McGill University: assistance in data collection.

Michelle Yuan, BSc., medical student at University of Alberta: assistance in data collection.

Conflict of Interest. None.

Statement of Authorship. LEK was responsible for data collection, statistical analysis, and wrote the original manuscript. ADR assisted with data collection and critically revised the manuscript for intellectual content. HMD assisted with data collection. NDS and MC provided expertise on the NIHTB-CB and critically revised the manuscript for intellectual content. TSF was responsible for the study design and inception and critically revised the manuscript for intellectual content.

References

- Weintraub S, Dikmen SS, Heaton RK, et al. Cognition assessment using the NIH Toolbox. *Neurology*. 2013;80:S54–S64. DOI [10.1212/WNL.0b013e3182872ded](https://doi.org/10.1212/WNL.0b013e3182872ded).
- Carlozzi NE, Goodnight S, Casaletto KB, et al. Validation of the NIH Toolbox in individuals with neurologic disorders. *Arch Clin Neuropsychol*. 2017;32:555–73. DOI [10.1093/arclin/axc020](https://doi.org/10.1093/arclin/axc020).
- Carlozzi NE, Tulskey DS, Wolf TJ, et al. Construct validity of the NIH Toolbox Cognition Battery in individuals with stroke. *Rehabil Psychol*. 2017;62:443–54. DOI [10.1037/rep0000195](https://doi.org/10.1037/rep0000195).

4. Heaton RK, Akshoomoff N, Tulsky D, et al. Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. *J Int Neuropsychol Soc.* 2014;20:588–98. DOI [10.1017/S1355617714000241](https://doi.org/10.1017/S1355617714000241).
5. Parsey CM, Bagger JE, Trittschuh EH, Hanson AJ. Utility of the iPad NIH toolbox cognition battery in a clinical trial of older adults. *J Am Geriatr Soc.* 2021;69:3519–28. DOI [10.1111/jgs.17382](https://doi.org/10.1111/jgs.17382).
6. Rebchuk AD, Deptuck HM, O'Neill ZR, Fawcett DS, Silverberg ND, Field TS. Validation of a novel telehealth administration protocol for the NIH toolbox-cognition battery. *Telemed e-Health.* 2019;25:237–42. DOI [10.1089/tmj.2018.0023](https://doi.org/10.1089/tmj.2018.0023).
7. Scott EP, Sorrell A, Benitez A. Psychometric properties of the NIH toolbox cognition battery in healthy older adults: reliability, validity, and agreement with standard neuropsychological tests. *J Int Neuropsychol Soc.* 2019;25:857–67. DOI [10.1017/S1355617719000614](https://doi.org/10.1017/S1355617719000614).
8. Roldán-Tapia L, García J, Cánovas R, León I. Cognitive reserve, age, and their relation to attentional and executive functions. *Appl Neuropsychol Adult.* 2012;19:2–8. DOI [10.1080/09084282.2011.595458](https://doi.org/10.1080/09084282.2011.595458).
9. Rebchuk AD, Kuzmuk LE, Deptuck HM, Silverberg ND, Field TS. Evaluating high-functioning young stroke survivors with cognitive complaints. *Can J Neurol Sci.* 2021;49:368–72. DOI [10.1017/cjn.2021.137](https://doi.org/10.1017/cjn.2021.137).
10. Jokinen H, Melkas S, Ylikoski R, et al. Post-stroke cognitive impairment is common even after successful clinical recovery. *Eur. J Neurol.* 2015;22:1288–94. DOI [10.1111/ene.12743](https://doi.org/10.1111/ene.12743).
11. Blackburn DJ, Bafadhel L, Randall M, Harkness KA. Cognitive screening in the acute stroke setting. *Age Ageing.* 2013;42:113–6. DOI [10.1093/ageing/afs116](https://doi.org/10.1093/ageing/afs116).
12. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 2007;39:175–91. DOI [10.3758/BF03193146](https://doi.org/10.3758/BF03193146).
13. Casaletto KB, Umlauf A, Beaumont J, et al. Demographically corrected normative standards for the English version of the NIH toolbox cognition battery. *J Int Neuropsychol Soc.* 2015;21:378–91. DOI [10.1017/S1355617715000351](https://doi.org/10.1017/S1355617715000351).
14. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol.* 2013;4:1–12. DOI [10.3389/fpsyg.2013.00863](https://doi.org/10.3389/fpsyg.2013.00863).
15. Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neurosci.* 2010;11:118–118. DOI [10.1186/1471-2202-11-118](https://doi.org/10.1186/1471-2202-11-118).
16. Tucker-Drob EM, de la Fuente J, Köhncke Y, Brandmaier AM, Nyberg L, Lindenberger U. A strong dependency between changes in fluid and crystallized abilities in human cognitive aging. *Sci Adv.* 2022;8:341. DOI [10.1126/sciadv.abj2422](https://doi.org/10.1126/sciadv.abj2422).
17. Thow ME, Summers MJ, Saunders NL, Summers JJ, Ritchie K, Vickers JC. Further education improves cognitive reserve and triggers improvement in selective cognitive functions in older adults: the Tasmanian healthy brain project. *Alzheimers Dement (Amst).* 2018;10:22–30. DOI [10.1016/j.dadm.2017.08.004](https://doi.org/10.1016/j.dadm.2017.08.004).