Paired Associates I and II; Wechsler Adult Intelligence Scale-IV Symbol Search, Digit Span, Coding, and Letter-Number Sequencing; Delis-Kaplan Executive Function System Color-Word Interference Test, Peabody Picture Vocabulary Test, Wechsler Individual Achievement Test-4 Spelling, and the Wisconsin Card Sorting Test), after which they completed MTB (pre-loaded on a study-provided smartphone) on their own. Internal consistency was evaluated using measure-appropriate indices (split-half reliability, Cronbach's alpha or IRT-based indices). Pearson correlation coefficients between MTB tests and measures of similar constructs were used to evaluate concurrent validity. For two tests with timing-dependent scores, Arrow Matching and Shape-Color Sorting, separate analyses were performed for iOS and Android devices. Sample II, with 1,120 English-speaking participants ages 18-90, was used to evaluate age-related change. Participants completed MTB measures remotely on their own smartphones, in a preset order, within a 14-day period. Spearman correlation coefficients, corrected for education, were calculated to evaluate relationships between age and test scores.
Results: Sample I participants were 67% female, 52% white, 99% non-Hispanic; average age=48 (SD= 17). Education was: < high school (1%); high school (55%); some college (21%); college (15%); graduate degree (8%). Internal consistency estimates ranged from 0.81 to 0.99. Pearson correlations between MTB and external measures ranged from 0.41 to 0.86 (all $p < .01$). Of the timed tests, only Shape-Color sorting showed significant score differences between Android and iOS devices.
Sample II was 57% female, 13% Hispanic, 72% white, mean age = 45 (SD = 21). Education distribution was: < high school (2%); high school (34%); some college (34%), college (20%); graduate degree (11%). Measures of executive function (r = -0.50; r=-0.57) and processing speed (r= -0.61) showed the expected negative correlation with age (all $p <0.001$). Negative correlations, although weaker, were also seen on measures of working memory (r=-0.2) and episodic memory (r=-0.2, r=-0.37; p.<.001). Vocabulary performance improved with age (r=0.4; p<.001), while spelling scores remained stable (r=0.09).
Conclusions: Initial studies support the validity and reliability of the first eight MTB cognitive measures in two diverse samples. MTB tests showed satisfactory construct validity, as demonstrated by the associations between MTB and well-established tests. Furthermore, most MTB measures correlated with age in the expected directions. Executive function, processing speed and memory typically decrease with age and this decrease was reflected in MTB test performance. In contrast, spelling and vocabulary, typically preserved as we age, did not decrease in our sample. Our results support the use of MTB in cognitive aging research.

Categories:
Assessment/Psychometrics/Methods (Adult)
Keyword 1: cognitive functioning
Keyword 2: assessment
Keyword 3: psychometrics
Correspondence: Cindy J. Nowinski, Northwestern University Feinberg School of Medicine

3 Mobile Toolbox: Enrollment of a Large Normative Sample Using the UCSF Brain Health Registry

Rachel L Nosheny[1,2], Monica R Camacho[2], Bernard Landavazo[2], Aaron J Kaat[3], Zahra Hosseinian[3], Richard C Gershon[3], Michael W Weiner[1,2]
[1]University of California San Francisco, San Francisco, CA, USA. [2]Veteran's Administration Advanced Research Center, San Francisco, CA, USA. [3]Northwestern University, Chicago, IL, USA

Objective: A critical need in the neuropsychology field is development and validation of efficient, scalable assessments of cognition. The Mobile Toolbox (MTB), a novel suite of mobile device-compatible, app-based cognitive assessments, was developed to address this need. The goals of this study were (1) To collect longitudinal normative data for the MTB assessments in a large, ethnoculturally and educationally diverse cohort; (2) To assess the feasibility and usability of remote assessment using MTB.
Participants and Methods: Participants were recruited from the UCSF Brain Health Registry (BHR), an online cohort (N>100,000) that collects longitudinal cognitive, functional,

https://doi.org/10.1017/S1355617723009736 Published online by Cambridge University Press

behavioral, and health data using online neuropsychological tests and self- and study-partner report surveys. BHR participants who opted to learning about additional research opportunities were sent automated email invitations to enroll in the MTB study. Those who indicated study interest were provided instructions within the BHR online portal for downloading the MTB app. All participants had the opportunity to complete a single baseline administration of MTB (Word Meaning, Sequences, Spelling, Arranging Pictures, Arrow Matching, Faces and Names, Shape-Color Sorting, Number Match). Those who completed the baseline assessment within three days were invited to continue into the longitudinal study, where they complete MTB assessments at a single, short-term timepoint (day 7, 14, or 21; study arms sequentially assigned), and then at 6-month intervals. Enrollment across demographic groups was monitored, and study invitations were sent to specific demographic groups, with the goal of enrolling a sample of 800 participants in the longitudinal study: equal distribution across eight, 10-year age bands (ages 18-80+); 60% with <16 years of education; 10% non-Latinx Black, 15% Latinx, and 5% non-White other ethnocultural identity.
Results: Between January-June 2022, 48,110 BHR participants were invited to the MTB study. Of those, 8294 (17%) expressed interest, 3401 (7%) completed the baseline assessment, 850 (1.8%) were assigned to the longitudinal study, and 782 (1.6%) completed a short-term longitudinal assessment. Study staff received 797 help tickets submitted by participants asking for email support to complete MTB. The baseline cohort had and average age of 64 years and an average of 16.6 years of education, 76.2% female, 2.1% non-Latinx Black, 7.1% Latinx, 86.8% non-Latinx White, and 4% from other ethnocultural groups. The longitudinal cohort had an average age of 62.3 years and an average of 16.1 years of education, 80% female, 2.8% non-Latinx Black, 8.5% Latinx, 83.5% non-Latinx White; and 5% other ethnocultural group. Compared to those invited to the study, those who enrolled in the longitudinal study were older, had higher educational attainment, and were more likely to be female and self-identify as non-Latinx White ($p<0.05$ for all).
Conclusions: Efficient enrollment and task completion of a large cohort in a novel, app-based mobile cognitive assessment is feasible in a completely remote setting. Most participants were able to complete MTB without individual support, indicating good usability. This approach can be scaled up to efficiently assess cognition in many research and healthcare settings. A remaining challenge is achieving robust ethnocultural and educational diversity.

Categories:
Assessment/Psychometrics/Methods (Adult)
Keyword 1: diversity
Keyword 2: normative data
Keyword 3: neuropsychological assessment
Correspondence: Rachel L. Nosheny
University of California San Francisco
rachel.nosheny@ucsf.edu

4 Continuum of Measurement: Reviewing the Advantages and Disadvantages of Self-Administered Remote Cognitive Tests and Their Examiner Administered Alternatives

Julie Hook, Cindy J. Nowinski, Richard Gershon
Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Objective: To describe advantages and disadvantages of using digital assessments remotely and in-person to inform clinical and research practice.
Participants and Methods: As part of a larger study,1120 adults completed a battery of remotely administered tests (Mobile Toolbox) and a subset of this sample completed examiner administered in-person testing (NIH Toolbox® Cognition Battery). Attention was given to making the sample reflective of the US 2020 Census during participant recruitment. Of the 1120 participants, the majority of the sample were female (57%) and Caucasian (72%) and had a mean age of 45 (SD = 21). In terms of education, equal percentages had high school (34%) or some college (34%).
Results: NIH Toolbox cognitive tests of processing speed, language, executive function, attention, and episodic memory were administered via a trained examiner and correlates of these tests were self-administered remotely via a smartphone. Using examples, we will show which aspects of cognitive assessment had the best correlations between remote self-administration and face-to-face examination and which had lower correlations.

https://doi.org/10.1017/S1355617723009736 Published online by Cambridge University Press