50 Remote Assessment has Minimal Effect on Test-Retest Reliability Among Older Adults with Essential Tremor

Sandra Rizer[1], Silvia Chapman[1], Jillian Joyce[1], Nikki Delgado[2], Margaret McGurn[2], Allison Powell[2], Daniella Iglesias Hernandez[2], Yian Gu[1], Elan D. Louis[2], Stephanie Cosentino[1]
[1]Columbia University Irving Medical Center, New York, NY, USA. [2]University of Texas Southwestern Medical Center, Dallas, Texas, USA

Objective: The COVID-19 pandemic increased utilization of remote assessment to allow clinicians and researchers to continue valuable work while maintaining quarantine guidelines. With guidelines relaxing, researchers have returned to in-person assessment. Information is needed regarding the effect of remote assessments on test-retest reliability. COGNET, a longitudinal study of cognition in participants with essential tremor, transitioned from in-person to remote assessments during the pandemic, and has now returned to in-person assessment. The current study investigates the extent to which remote assessment affected test-retest reliability across a range of neuropsychological assessments administered in COGNET.
Participants and Methods: Participants included 27 older adults enrolled in COGNET (mean age=75.0 (9.1), education=16.2 (2.6), 67% female, and 100% white). Memory tests included: California Verbal Learning Test II, Logical Memory subtest of the Wechsler Memory Scales – Revised, and Verbal Paired? Associates. Executive function tests included: Digit Span Backwards and the Delis-Kaplan Executive Function System subtests of Verbal Fluency, Sorting, and Color-Word. Attention tests included Oral Symbol Digit Modalities Test and Digit Span Forward. Language was assessed with the Boston Naming Test. Intraclass correlation coefficients (ICCs) were calculated to examine test-retest reliability of In-Person to In-Person visits (P-P), and combination visits (e.g., In-Person to Remote (P-R), and Remote to In-Person (R-P)). Following Koo & Li (2016), ICCs were interpreted as: >.90 excellent, .75-.90 good, .50-.74 moderate, and <.50 poor reliability. The Feldt approach was used to compare ICCs from P-P visits against ICCs calculated for combination visits (P-R or R-P), with the test statistic compared to an F distribution.
Results: ICCs for person-to-person assessment ranged from .51 to .89. Memory test ICCs ranged from moderate to good (.51 to .80). Executive function test ICCs ranged from moderate to good (.55 to .89). The attention domain had moderate ICCs (.67 - .68). Language ICC was moderate (.70). ICCs for person-to-remote assessment ranged from .42 to .89. Memory tests ranged from moderate to good ICCs (.59 to .83). Executive function tests ranged from poor to good ICCs (.42 to .89). Attention ICCs were moderate to good (.55 to .79). The Language ICC was moderate (.72). ICCs for remote-to-person ranged from .48 to 86. Memory ICCs ranged from moderate to good (.59 to .86). Executive function ICCs ranged from poor to good (.48 to .83). Attention ICCs were moderate to good (.56 to .79). The Language ICC was good (.78). The only test for which an ICC from a combination visit was significantly lower than a person to person visit was Digit Span Backwards.
Conclusions: Test-retest reliability was moderate or better for all P-P assessments, consistent with the known psychometrics of these tests. Only one test of executive function showed lower reliability when remote assessment was introduced. From a broad standpoint, current results suggest that remote administration of neuropsychological tests can be used as a reliable substitute for in-person assessment for many measures, and suggest that caution be used when interpreting any change in Digit Span Backwards across person and remote assessments.

51 Importance of Using Baseline Verbal Abilities When Interpreting the MOCA Test Performance