HHS Public Access

Author manuscript

Clin Neuropsychol. Author manuscript; available in PMC 2022 March 07.

Published in final edited form as:

Clin Neuropsychol. 2020 ; 34(7-8): 1411–1452. doi:10.1080/13854046.2020.1769192.

Validity of Teleneuropsychology for Older Adults in Response to COVID-19: A Systematic and Critical Review

David E. Marra, PhD[1,2], **Kristin M. Hamlet, PhD**[1], **Russell M. Bauer, PhD., ABPP-CN**[1,3,4], **Dawn Bowers, PhD., ABPP-CN**[1,3]

[1]Department of Clinical and Health Psychology, University of Florida, Gainesville, FL

[2]McKnight Brain Institute, University of Florida, Gainesville, FL

[3]Department of Neurology, University of Florida, Gainesville, FL

[4]Brain Rehabilitation Research Center, Malcolm Randall VAMC, Gainesville, FL

Abstract

Objective: Due to the recent COVID-19 pandemic, the field of neuropsychology must rapidly evolve to incorporate assessments delivered via telehealth, or teleneuropsychology (TNP). Given the increasing demand to deliver services electronically due to public health concerns, it is important to review available TNP validity studies. This systematic review builds upon the work of Brearly and colleagues' (2017) meta-analysis and provides an updated review of the literature, with special emphasis on test-level validity data.

Method: Using similar methodology as Brearly and colleagues (2017) three internet databases (PubMed, EBSCOhost, PsycINFO) were searched for relevant articles published since 2016. Studies with older adults (aged 65+) who underwent face-to-face and TNP assessments in a counterbalanced cross-over design were included. After review, 10 articles were retained. Combined with 9 articles from Brearly's (2017) analysis, a total of 19 studies were included in the systematic review.

Results: Retained studies included samples from 5 different countries, various ethnic/cultural backgrounds, and diverse diagnostic populations. Test-level analysis suggests there are cognitive screeners (MMSE, MoCA), language tests (BNT, Letter Fluency), attention/working memory tasks (Digit Span Total), and memory tests (HVLT-R) with strong support for TNP validity. Other measures are promising but lack sufficient support at this time. Few TNP studies have done in-home assessments and most studies rely on a PC or laptop.

Conclusions: Overall, there appears to be good support for TNP assessments in older adults. Challenges to TNP in the current climate are discussed. Finally, a provisional outline of viable TNP procedures used in our clinic is provided.

Keywords

Telehealth; Teleneuropsychology; Systematic Review; Validity Study

Corresponding Author: Send correspondences to David Marra dmarra3@phhp.ufl.edu. McKnight Brain Institute 1149 Newell Drive, Room 3110, Gainesville, FL 32610.

On March 11, 2020, the World Health Organization characterized the outbreak caused by the novel COVID-19 virus as a pandemic (Adhanom Ghebreyesus, T., 2020). COVID-19 is a respiratory disease that can cause mild to severe illness. As of 26 April 2020, there were 2,804,796 cases worldwide with 193,710 deaths; in the United States there were 899,281 confirmed cases and 38,509 deaths (World Health Organization, 2020). Due to the highly contagious nature of the virus, it is spreading widely throughout the world and exhibiting "hotspots" conforming to population density (e.g., New York City). According to the CDC, adults 65 and older are particularly at risk of severe illness from COVID-19 and have a higher mortality rate than their younger counterparts (Centers for Disease Control and Prevention, 2020). Of the adults aged 65 and older in the U.S. with confirmed cases of COVID-19, 31–59% required hospitalization and11–31% required admission to an intensive care unit. Eight out of ten deaths due to COVID-19 in the U.S. have been older adults. Thus, world-wide efforts are underway to protect the public and "flatten the curve" of COVID-19 incidence. Such efforts include social distancing, self-quarantine, and "stay at home" orders. Thus, the utilization of telehealth has become critical to allow access to medical care during this pandemic.

In response to the increased demand for health care service delivery via telehealth, Medicare has relaxed some of the pre-existing regulations for telehealth services and will reimburse at the same dollar-amount as in-person visits (Coronavirus Preparedness and Response Supplemental Appropriations Act, 2020). Similarly, the Department of Health and Human Services (HHS) has relaxed HIPPA privacy laws such that a provider, who is practicing in good faith, can use any non-public facing remote communication production that is available (Office for Civil Rights, 2020).

Diagnostic and interventional telehealth services have been well-established for age-related cognitive decline and dementia, especially in underserved and rural communities. A recent review found good support for telehealth in the assessment and management of patients with Parkinson's disease (PD) and Alzheimer's disease (AD) (Adams, Myers, Waddell, Spear, & Schneider, 2020). Similarly, a recent systematic review found almost perfect correspondence between in-person and telehealth assessments when diagnosing AD via clinical interview. Furthermore, telehealth may also be useful for early detection of MCI and preclinical dementia (Costanzo et al., 2020).

Despite its growing support, neuropsychological assessment delivered via telehealth (i.e., teleneuropsychology) has not been utilized by the vast majority of practitioners. This may largely be due to the previous lack of reimbursement from Medicare and private insurances. Additionally, there are some challenges and criticisms of teleneuropsychology (TNP) assessments. Such challenges include limited access to or familiarity with technological services (i.e., high-speed internet, web camera), inability to perform "hands-on" portions of an assessment, and reduced opportunities for behavioral observations due to camera angles (Barton, Morris, Rothlind, & Yaffe, 2011; Brearly et al., 2007; Harrell, Wilkins, Connor, & Chodosh, 2014; Parikh et al., 2013; Turner, Horner, Vankirk, Myrick, & Tuerk, 2012). Most studies utilizing TNP assessments take place in a satellite clinic, where a technician can set up and configure the equipment and necessary test stimuli. However, given a desire to abide by appropriate social distancing practices, TNP assessments at satellite clinics

may not be possible; rather the in-home model of TNP assessments may be favored. Thus, in-home TNP assessments would occur in a non-controlled environment where there may be more interruptions and there is no control over test material created by the patient, raising concerns for test security. In addition, there is a concern that normative data, derived from standardized test procedures, may not be appropriate for TNP evaluations (Brearly et al., 2017). Given all of this, there are some clinicians who believe TNP evaluations may be unethical, especially for high-stakes evaluations (e.g., forensic and competency evaluations). Finally, there is the ethical consideration of Justice, in that it is unclear whether all patients can be equally served via TNP due to lack of appropriate computer/internet connections (low SES) or disability (visually, hearing impaired), potentially exacerbating existing problems with healthcare delivery system.

While there are some notable concerns to TNP evaluations, proponents point to several benefits. First, there is generally positive feedback from patients and caregivers regarding these services (Barton et al., 2011; Harrell et al.,, 2014; Parikh et al., 2013; Turner et al., 2012). TNP assessments can also reach a wider population of individuals who may have restricted mobility or live long distances from the clinic (Brearly et al., 2017). Regarding the on-going pandemic, private insurances and Medicare are temporarily reimbursing telehealth visits at the same dollar-amount as in-person services (Coronavirus Preparedness and Response Supplemental Appropriations Act, 2020), allowing neuropsychologists to continue providing services to patients from their home, without the added risk of virus exposure. TNP may also facilitate connectedness with patients, many of whom need services and interpersonal contact. Finally, TNP may facilitate other medical treatment when stay-in-place orders are lifted.

With regard to TNP validity, some validity studies showed subtle differences in task-performance when comparing face-to-face (FTF) with TNP assessments (Cullum, Weiner, Gehrmann, & Hynan, 2006; Grosch, Weiner, Hynan, Shore, & Cullum, 2015; Hildebrand, Chow, Williams, Nelson, & Wass, 2004; Wadsworth et al., 2018; Wadsworth et al., 2016); though many other studies showed no such performance differences (Ciemins, Holloway, Coon, McClosky-Armstrong, & Min, 2009; DeYoung & Shenal, 2019; Galusha-Glasscock, Horton, Weiner, & Cullum, 2015; McEachern, Kirk, Morgan, Crossley, & Henry, 2008; Menon et al., 2001; Turkstra, Quinn-Padron, Johnson, Workinger, & Antoniotti, 2012; Vahia et al., 2015; Vestal, Smith-Olinde, Hicks, Hutton, & Hart, 2006). To evaluate potential performance differences between TNP and FTF assessments, Brearly and colleagues (2017) conducted a meta-analysis of 12 studies published between 1997 and 2016 (see Table 1). The included studies were counter-balanced cross-over designs (FTF, virtual) with adult patient samples (>17 years of age). Studies were excluded if active involvement of an assistant was required during testing (e.g., more than just showing a participant how to adjust volume).

Across the 12 included studies, a total sample of 497 study participants and patients were included. The overall effect size distinguishing TNP from FTF performance was small and non-significant ($g = -0.03$; SE = 0.03; 95% CI [−0.08, 0.02], $p = .253$). Across all 79 scores from included studies, 26 mean scores were higher for the videoconference condition (32.91%), 48 mean scores were higher for the FTF condition (60.76%), and five mean scores

were exactly the same in both conditions (6.33%). Further analysis showed a small, but significant effect size (g = −0.10; SE = 0.03; 95% CI [−0.16, −0.04], p < .001) for timed tests or tests where a disruption of stimulus presentation may affect test results (e.g., digit span, list-learning tests), with TNP performance approximately 1/10 of a SD lower than FTF testing performance. A similar magnitude of difference was found for the BNT-15 items (g = −.12; SE = .03, p <.001). Finally, a moderator analysis showed that there was no difference in FTF vs. virtual performance for adults aged 65–75 (g = 0.00, SE = .01, p = .162). Further moderator analyses were not interpreted due to significant heterogeneity between sub-groups. The authors concluded "the current findings did not reveal a clear trend towards inferior performance when tests were administered via videoconference. Consistent differences were found for only one test (BNT-15) and the effect size was small" (Brearly et al., 2017, pg. 183).

The meta-analysis conducted by Brearly and colleagues (2017) was a *critical* first step to demonstrate the relative validity and utility of TNP. While this review was quite useful in objectively and quantitatively demonstrating the utility of various neurocognitive assessments in the TNP environment, it lacked a qualitative analysis of the available validity data for each assessment. As such, it may be difficult to critically appraise the available validity evidence (e.g., sample size, demographic composition) of *each test* when selecting a test battery for TNP.

The current project is a limited systematic review that builds on the important work of Brearly and colleagues (2017). Using similar methodology, the present review provides an updated qualitive analysis and test-level data from each validity study. Given older adult's particular susceptibility to COVID-19, we limited our analysis to studies of adults aged 65 and older. While test-level validity data for various measures delivered via TNP was the primary objective of this systematic review, we also conducted a critical (non-systematic) review of the modality in which TNP services were delivered as well as an appraisal of validity studies that included ethnic minority populations.

Methods

This review was conducted in accordance with Preferred Reporting Items for Systematic Review and Meta-analyses guidelines (PRISMA; Moher, Liberati, Tetzlaff, & Altman, 2009) and was pre-registered with PROSPERO, an international prospective register of systematic reviews (PROSPERO ID: 175521).

Article Search and Selection

For consistency, article search and selection largely mirrored the methodology used by Brearly and colleagues (see Brearly et al., 2017 for full details). Briefly, the same three internet databases (PubMed, EBSCO (PsycINFO), ProQuest) were searched for relevant articles using the terms, "(tele OR remote OR video OR cyber) AND cognitive AND (testing OR assessment OR evaluation).[1]" Diverging from their methodology, additional specifiers

[1]Full search term from PubMed: ((((tele OR remote OR video OR cyber))) AND cognitive) AND (testing OR assessment OR evaluation) AND ("2016/01/01"[PDat] : "2020/03/21"[PDat]) AND aged[MeSH

were included to align with the aims of the present systematic review and identify studies involving older adults (ages 65+). In addition, the article search was limited to studies published after 1/1/2016, the endpoint of Brearly and colleagues' (2017) article search.

Articles were included if the average age of the study sample was 65 or greater and neuropsychological assessments were conducted with a counter-balanced cross-over design where participants were assessed FTF and via videoconference. Articles were excluded if inferential statistics for test-level data were not included, if participants required significant in-person assistance from a technician or test administrator, or if studies "utilized proprietary software or hardware specifically designed for test administration (e.g., touchscreen kiosks, mobile applications)." (Brearly et al., 2017, pg. 176). Studies from Brearly's analysis were also included in the following qualitative analysis when the average age of participants was 65+.

Article extraction took place on 3/21/2020. A total of 591 articles were extracted across the three internet databases, with 532 remaining after duplicates were removed. Nine articles were identified for review from informal searches and from the reference sections of published articles. Titles and abstracts were reviewed by the primary author for potential inclusion. The initial screening process was managed by the open-sourced software, abstrackr (Wallace, Small, Brodley, Lau, & Trikalinos, 2012). 54 of 532 (10%) of abstracts from the initial screening were double-coded to by the primary author to establish reliability, which was perfect ($\kappa = 1.00$, $p < .0001$). After initial review, 24 articles were selected for full-text review. Seven unique articles were included after full-text review (see Figure 1)

Three articles did not meet full inclusionary/exclusionary criteria as outlined by Brearly (2017) but were included because the authors felt they were informative to the systematic review. In two of these studies (Abdolahi et al., 2016; Stillerova, Liddle, Gustafsson, Lamont, & Silburn, 2016) the cross-over design was not counterbalanced. That is, the participants did FTF assessments followed by remote assessments. However, these studies were retained because the study sample included participants with movement disorders and, uniquely, the assessments were conducted at the participants' homes. A third study (Vahia et al., 2015) provided inferential statistics for the overall analyses, but did not provide test-level data and inferential statistics. However, this study was retained because it included an exclusively Hispanic sample, with tests administered in Spanish.

After full text review, 10 studies published since 2016 were retained for systematic review. Nine articles from Brearly and colleagues (2017) analysis were also included for qualitative review. In sum, a total of a total 19 articles were included in the systematic review.

Study Quality and Risk of Bias

PRISMA guidelines suggest an analysis of study quality and risk of bias. However, checklists to assess these domains, such as the Cochrane Review Checklist, are ill-equipped for assessing bias in cross-study designs (Brearly et al., 2017; Ding et al., 2015). An alternate checklist for cross-over designs was proposed by Ding and colleagues (2015). However, this checklist is problematic as neither clinicians nor study participants can be blinded to condition (FTF vs. TNP). Based on items that could be applied to the studies

of this review, almost all studies were deemed to be of moderate to high quality based on the requirement that all included studies were counter-balanced cross-over designs with all outcomes reported. The three studies described immediately above (Abdolahi et al., 2016; Stillerova et al., 2016; Vahia et al., 2015) were of lower study quality due to a lack of counter-balance in the cross-over design and failure to report all outcomes.

Publication Bias

Publication bias arises when studies with larger effect sizes are more likely to get published, whereas studies with null findings or smaller effect sizes are less likely to be published (i.e., file-drawer effect). However, as noted by Brearly and colleagues (2017), the risk of publication bias is likely low as authors in this field are more likely to publish articles in which effect sizes are small. Nonetheless, a quantitative analysis of the effect sizes from Brearly and colleagues' (2017) study (9 of the 19 articles from the current review) showed no evidence of bias (symmetry around the funnel plot, Kendell's tau $b = -.227$, $p = .304$). Formal assessment of publication bias for the present review was not conducted (e.g., funnel plot, Egger's Test) as effect sizes were not calculated in this study but were qualitatively reviewed, only when provided by study authors.

Assessing Teleneuropsychology Validity

The validity of TNP was assessed via several facets. First, through the authors' report of mean performance differences across testing environments (FTF vs. TNP). Effect sizes (Cohen's d, Hedge's *g*, Pearson *r*), which measures the standardized difference between two means, were interpreted according to conventions established by Cohen (1988) (i.e., Cohen's *d* and Hedge's *g* of .20 = small, .50 = medium, .80 = large; Pearson's *r* of .10 = small, .30 = medium, .50 = large). Absolute Intraclass correlation (ICC), a metric of test-retest reliability (Koo & Li, 2016), was used to describe validity of TNP testing performance relative to FTF performance. ICC was interpreted based on conventions established by Cicchetti (1994) (i.e., ICC 0-.39 = poor, .40-.59 = fair, .60-.74 = good, .75–1.00 = excellent). Similar to ICC, Cohen's kappa measures reliability, but for categorical items and corrects for agreement that may have occurred by chance (Cohen, 1988). Interpretation of Cohen's kappa was based on conventions established by Cohen (1988) (i.e., kappa 0-.20 = none, .21-.39 = minimal; .40=.59 = weak, .60 - .79 = moderate, .80 - .90 = strong, > .90 = almost perfect). Finally, the Bland-Altman plot, a calculation of the mean difference between two assessment methods, is a method of assessing bias (Bland & Altman, 1986). This method yields a 95% Limits of Agreement. If the 95% Limits of Agreement includes 0, then there is no evidence of systematic bias favoring performance in FTF or TNP.

Each measure was qualitatively judged to have either strong, moderate, or limited/ insufficient evidence of TNP validity based on tiered review of available evidence of: 1) the number of available validity studies for each measure and the between-study agreement; and 2) the sample size and diagnostic characteristics of the validity studies. That is, a measure was considered having strong TNP validity if there were multiple validity studies - some with large sample sizes and diagnostically diverse patient populations – that showed relatively good between-study agreement. A measure was considered to have moderate TNP validity evidence if there were multiple validity studies with some between-study

variability or a few large studies with diagnostically diverse patient populations that showed good between-study variability. Finally, studies were considered to have limited/insufficient evidence of TNP validity if there were few validity studies with small sample sizes/lack of diagnostic diversity or extreme between-study variability.

Results

Study Characteristics

Diagnostic Groups.—There is a wide range of diagnostic samples represented in the 19 TNP validity studies. Diagnostic groups included participants with movement disorders (Abdolahi et al., 2016; Stillerova et al., 2016), stroke/cerebrovascular accident (Chapman et al., 2019), psychiatric diagnoses (Grosch et al., 2015), and mild cognitive impairment (MCI) or Alzheimer's disease (AD) (Carotenuto et al., 2018; Cullum et al., 2006; Loh, Donaldson, Flicker, Maher, & Goldswain, 2007; Vestal et al., 2006). Other samples were mixed with healthy controls (HC) and patients with psychiatric conditions or memory disorders (Cullum, Hynan, Grosch, Parikh, & Weiner, 2014; Lindauer et al., 2017; Loh et al., 2004; Montani et al., 1997; Wadsworth et al., 2018; Wadsworth et al., 2016). Of the 930 study participants across the 19 validity studies, n = 410 (44.09%) were healthy controls, n = 359 (38.60%) were patients with memory disorders, n = 28 (3.01%) were patients with movement disorders, n = 78 (8.39%) were patients with stroke/cerebrovascular accident, n = 30 (3.22%) were psychiatric patients, and n = 34 (3.65%) were patients from mixed clinical groups with no further diagnostic differentiation.

Diagnosis was a potential confound to TNP validity in only one study; Abdolahi and colleagues (2016) found the psychometric properties of the Montreal Cognitive Assessment (MoCA) to be relatively poor in a group of participants with PD (ICC = .37; Cronbach's alpha = .54, Pearson r = .37) compared to a group with Huntington's disease (HD) (ICC = .65; Cronbach's alpha = .79; Pearson's r = .65). However, the sample size of this study was relatively small (n = 8 PD; n = 9 HD), there were differences in follow-up assessments (7 months PD vs. 3 months HD), and the cross-over design was not counter-balanced. Finally, the authors did not control for the assessment time-of-day, potentially introducing variability in response to dopaminergic medications. In contrast, a study with similar methodology and PD participants found relatively good reliability, with a median difference score of only 2 (IQR = 1.0–2.5) out of 30 points (Stillerova et al., 2016). Thus, with these minor exceptions, diagnosis does not seem to affect TNP validity (see Disease Severity).

Cultural and Racial Groups.—Seven of the included studies were conducted in countries outside of the United States. One study was conducted in Italy (Carotenuto et al., 2018), three in Australia (Loh et al., 2007; Loh et al., 2004; Stillerova et al., 2016), one in Korea (Park, Jeon, Lee, Cho, & Park, 2017), one in Canada (Hildebrand et al., 2004), and one in Japan (Yoshida et al., 2019).

Of the studies conducted in the United States, there was an underrepresentation of ethnic minorities. However, two studies included samples that were 100% ethnic minorities. Vahia and colleagues (2015) conducted a study with monolingual and bilingual Hispanics with testing completed exclusively in Spanish. Wadsworth and colleagues (2016) conducted a

validity study with a sample of American Indians. Of the studies with mixed demographic compositions, non-Hispanic Caucasians were overrepresented, with little representation from Hispanics, African Americans, or ethnic minorities. African Americans had little or no representation in these validity studies.

Teleneuropsychology Test Validity[2]

Cognitive Screeners.—The brief cognitive screener with the most support is the Mini Mental State Examination (MMSE), with nine of the ten studies reporting no mean differences in test scores when comparing TNP to FTF administration (See Table 2). In the one study where mean differences were observed, the effect size was small-medium ($g = -.40$, $p < .001$), but a strong correlation was observed between scores in the two testing modalities ($r = .95$) (Montani et al., 1997). Psychometrics for reliability were generally excellent (ICC ranged from .42 −.92; Pearson r ranged from .90-.95). The strongest support comes from Cullum and colleagues (2014) which had a large sample of MCI/AD patients (n = 83) and healthy controls (n = 199) and showed excellent reliability (ICC = .798). The MMSE was also valid during longitudinal assessments. Carotenuto and colleagues (2018) conducted serial MMSE assessments at baseline, 6, 12, 18, and 24 months in a sample of 28 Italian patients with AD. Across they whole AD group, they found no mean differences in performance across testing modalities at any time-point. Taken together, it appears the MMSE is a valid telehealth measure for screening cognitive status across different clinical populations and also has utility as a longitudinal assessment measure.

The Montreal Cognitive Assessment (MoCA) was used in four validity studies, though as described above, two studies (Abdolahi et al., 2018; Stillerova et al., 2016) did not use a counter-balanced cross-over design. Nonetheless, the psychometrics appear sound with strong reliability metrics (ICC range from .59-.93) and no study finding mean TNP vs. FTF differences. In a study of 48 stroke survivors, neither age, computer literacy, nor self-reported anxiety/depression predicted differences in scores between testing conditions (Chapman et al., 2019). Thus, individual factors may not account for TNP vs. FTF differences. Taken together, while only two of the four validity studies were counter-balanced cross-over designs, there appears to be good validity for using the MoCA in TNP assessments.

Two validity studies utilized the Alzheimer's Disease Assessment Scale – Cognitive Subscale (ADAS-cog), one consisting of a mixed sample of AD/MCI/HC patients (Yoshida et al., 2019) and one involving a two-year longitudinal study of AD patients from Italy (Carotenuto et al., 2018). Yoshida and colleagues showed excellent reliability metrics (ICC = .86). Similarly, Carotenuto (2018) found no mean differences in performance across any time point (baseline, 12, 18, and 24 months) for the whole group of AD patients. Although there are only two validity studies, the sample size from one of the studies was relatively large (N = 73) and another showed relatively good validity for mild-to-moderate AD patients across five different assessment periods.

[2]Data from Tables 2 – 8 were made available to the public on the Inter Organizational Practice Committee website prior to manuscript publication in service to other practitioners (https://iopc.squarespace.com/teleneuropsychology-research).

A single study utilized the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Galusha-Glassock et al., 2016). In this mixed sample of MCI, AD, and HC (N = 18), average RBANS Total score and all other index scores were statistically similar across testing environments. The ICC for the Total score was excellent (ICC = .88). Reliability of the visuospatial/constructional index score was fair (ICC = .59), whereas the ICC for every other index score was excellent (ICC range .75-.90). Notably, however, the record form for the Coding sub-test was left with the patient, who was assessed in a non-adjacent room of the same facility. Therefore, giving the full RBANS may not be practical when the patient is assessed from home unless materials are mailed to the patient in advance of the appointment. Given the small sample size of the single validity study and the difficulty that would accompany getting patients the appropriate record forms, there is limited support for the validity of the RBANS for TNP assessments. Rather, it may be beneficial to utilize select subtests from the RBANS (e.g., Line Orientation) to supplement a neurocognitive assessment.

Intelligence.—Only one study (see Table 3) assessed intellectual functioning using the Matrix Reasoning and Vocabulary subtests from the Wechsler Adulty Intelligence Scale – 3rd edition (WAIS-III) (Hildebrand et al., 2004). This relatively small sample (N = 29) consisted of a HC population from Canada. There were no TNP vs. FTF differences in mean scores for either subtest. Furthermore, the limits of agreement did not suggest performance bias in either domain (95% Limits of Agreement MR: −4.56 – 6.08; 95% Limits of Agreement Vocabulary: −3.07–3.13). While measures of intellectual functioning are an instrumental part of a neuropsychological assessment, research is lacking in this domain and there is presently limited support for telehealth validity of such measures, especially in clinical samples.

Attention/Working Memory.—Six unique validity studies utilized the Digit Span task (Digit Span Forwards, Digit Span Backwards, Digit Span Total; see Table 4). These studies included minority samples (American Indians, Hispanics) and included different diagnostic groups (MCI, AD, HC, Psychiatric samples). For Digit Span Forward, three of the four studies found no TNP vs. FTF difference in mean performance. In the one study that did find significant differences (Wadsworth et al., 2016), a small effect size favoring FTF performance was reported. Of similar concern, the largest validity study (Cullum et al., 2014) consisting of 202 MCI, AD, and HC patients, reported only fair validity statistics (ICC = .590). In this study, the authors used an alternative version of the digit span task to reduce practice effects, though they did not indicate which versions were used. They did acknowledge, “it is possible that our choice of alternate digit strings resulted in lower correlations, and other versions (e.g., WAIS-4, RBANS) may show higher correlations.” (Cullum et al., 2014, pg. 6). Alternatively, another large study (N = 197 AD, MCI, HC) found no TNP v. FTF main effects after controlling for age, education, gender, and depression) with very small effect sizes (d = .007) for the clinical group (Wadsworth et al., 2018). Thus, there appears to be moderate evidence of validity for utilizing Digit Span Forward in TNP assessments.

A very similar pattern emerges for studies that utilized Digit Span Backwards. While no study found significant TNP v. FTF differences in mean performance, the largest study (Cullum et al., 2014) observed only fair validity metrics (ICC = .545); whereas, another large study with a mixed group (Wadsworth et al., 2018) found no such effects after controlling for age, education, gender, and depression scores, and very small effect sizes for the clinical group (d = .048). Again, there is moderate evidence of validity for TNP assessments.

Finally, for the two studies that reported Digit Span Total (Cullum et al., 2006; Grosch et al., 2015), reliability metrics were good and excellent (ICC = .72 and .78, respectively). Though the samples were small, they were mixed samples of AD, MCI, and geropsychiatric patients. Thus, there appears to be good validity evidence for using Digit Span Total in TNP.

One study using a HC sample of individuals from Canada (Hildebrand et al., 2004) used the Brief Test of Attention (BTA) and the authors found no mean differences in scores resulting from TNP vs. FTF assessments. There also was no evidence of bias towards a particular testing modality (95% Limits of Agreement: −5.09 – 6.95). Given the small sample size with no clinical patients, there is limited validity evidence for the BTA to be used in TNP assessments of a clinical population.

Processing Speed.—Only one validity study used a measure of processing speed (see Table 5), Oral Trails A (Wadsworth et al., 2016). This mixed-sample study (AD, MCI, HC) showed a significant difference in completion time, with a better or faster performance in person. The authors maintain that the difference in performance was small (Mean Time = 8.9 (SD = 2.4) vs. 11.1 (SD = 3.0) and not clinically meaningful as they fell within the normal range of test-retest reliability. Similarly, the validity metrics for this test were excellent (ICC = .83). Taken together, there is some support for the validity of Oral Trails A in TNP assessments.

Language.—A small study of ten Veterans referred for a memory disorders evaluation was assessed using the full 60-item Boston Naming Test (BNT), which found no difference in same-day performance between TNP and FTF assessment modalities (Vestal et al., 2006). The 15-item BNT (BNT-15) was used in four studies with medium to large samples (N range from 33–202) of mixed clinical and healthy samples, which showed excellent reliability metrics (ICC ranged from .812 to .930). A significant mean TNP vs. FTF performance difference was found in only one of the four studies (Wadsworth et al., 2016), but the effect size was small in favor of FTF assessment (g = −0.15, p < .001). Overall, there appears to be good support for the validity of the BNT in TNP assessments (see Table 6).

Vahia and colleagues (2015) administered the Ponton-Satz Spanish Naming Test to 22 Spanish-speaking Hispanics who were referred for a memory disorders evaluation by their psychiatrist. Using a mixed-effects model, the authors found no significant TNP vs. FTF performance differences (though means and standard deviations were not provided). Taken together, there is some evidence of validity for TNP with this population.

With regards to letter fluency, there are seven validity studies with sample sizes ranging from small (N = 10) to large (N = 202), consisting of multi-ethnic demographics (American Indians, Spanish-speaking Hispanics), with various clinical samples (MCI, AD, HC, psychiatric). No mean differences between TNP and FTF evaluations were reported in any study and validity metrics were excellent (ICC = .83 to .93). There appears to be strong support for the validity of letter fluency in TNP assessments.

Category fluency results were slightly more variable. There were five validity studies that administered category fluency with similar demographic and clinical compositions as letter fluency. However, one study (Wadsworth et al., 2018) found a small, but significant difference in testing modality (d = .184 for MCI/AD group). Similarly, validity metrics were only fair-to-good (ICC ranged from .58 - .74). This may be due to the use of a single trial semantic/category fluency measure (e.g., Animals), which likely creates more variability performance. Thus, there is moderate validity for using category fluency as part of a TNP assessment, though it may be beneficial to use a category fluency test with multiple trials (e.g., Animals, Vegetables, Fruits).

As part of their language evaluation, Vestal and colleagues (2006) administered the Token Test, Picture Description, and Aural Comprehension of Words and Phrases to 10 Veterans who were referred for memory disorders evaluations. There were no significant TNP vs. FTF performance differences. Notably, the Veterans were given the tokens and a template to re-organize the stimuli for the Token test, making it unlikely to be useful in telehealth evaluations unless examiners can find creative ways to provide distant examinees with appropriate stimulus materials in advance of the assessment. Given the small sample size of this single study, there is insufficient evidence for the validity of these measures in TNP assessments.

Memory.[3]**—**Five studies examined the validity of the Hopkins Verbal Learning Test – Revised (HVLT-R). The sample size of these studies ranged from medium (N = 22) to large (N = 202), consisting of multi-ethnic demographics (American Indians, Spanish-speaking Hispanics), with various clinical samples (MCI, AD, HC, psychiatric). For HVLT-R Immediate Recall Total, only one study found significant TNP v. FTF performance differences (Cullum et al., 2014), though the effect size was small (g = .13, p = .004). Validity metrics were excellent (ICC = .77 - .88). Three of the studies mentioned above also examined HVLT-R Delayed Recall, which found no significant differences in performances; validity metrics were good-to-excellent (ICC = .61 & .90). Taken together, the HVLT-R has strong support for validity in TNP assessments (see Table 7).

One study used the Brief Visuospatial Memory Test – Revised (BVMT-R), which included a sample of 22 Spanish-speaking Hispanics. A mixed-effects model showed no significant TNP v. FTF differences, though mean scores were not provided. Taken together, there appears to be some evidence of validity for using BVMT-R in TNP evaluations.

[3]Hildebrand and colleagues (2004) also administered the Rey Auditory Verbal Learning Test as part of their validity study. However, in personal correspondence with Brearly and colleagues (2017), the authors identified potential problems with the RAVLT data and was excluded from Brearly's quantitative analysis (Brearly et al., 2017, pg. 179). Therefore, the RAVLT was not included in this systematic review.

Executive Functioning.—In the review of available studies, many traditional measures of executive functioning (e.g., Wisconsin Card Sort Test, Trail Making Test B, Stroop) were not assessed. Instead, the most widely used measure of "executive functioning" was the Clock Drawing Test, which involves multi-componential processes including visuospatial skills, language, as well as executive skills such as planning and inhibitory control.

The Clock Drawing Test was used in eight validity studies, ranging from small (N = 8) to large (N = 202) sample sizes, with different ethnic compositions (American Indian, Hispanic) and clinical populations (MCI, AD, HC, rehabilitation, psychiatric). While no study reported significant differences in mean performance between testing conditions, there was variability in findings and validity metrics. For example, two studies reported large, but non-significant differences between TNP and FTF evaluations (Hildebrand et al., 2004; Montani et al., 1997). Furthermore, validity metrics ranged for poor-to-good (ICC range .42 - .71) with only moderate reliability (kappa = .48). However, the studies that reported poorer validity metrics tended to be much smaller than the others. For example, Grosch (2015) found poor validity metrics (ICC = .42) with a sample of only 8 patients seen in a VA geropsychiatry clinic. The three largest studies (Cullum, 2014; Wadsworth, 2016, 2018) reported no significant TNP v. FTF differences and good validity metrics (ICC = .65 and .71). Given the variability in findings and only moderate validity metrics, there appears to be only moderate evidence for the reliability of the Clock Drawing Test in TNP (see Table 8).

Disease Severity

Two studies discussed validity metrics for TNP based on disease severity. Corotenuto and colleagues (2018) assessed AD patients with the MMSE and ADAS-Cog at baseline, 6 12, 18, and 24 months. When separated by severity (characterized as mild, moderate, and severe), MMSE videoconference performance for the severe AD patients (MMSE 15–17) was worse than FTF performance at baseline and 24 months; whereas, the mild (MMSE 21–24) and moderate (MMSE 18–20) AD patients did not show performance differences at any time. Similarly, they found that patients with severe AD had significantly higher (worse) ADAS-Cog scores during teleconference evaluation, whereas there were no performance differences in the mild and moderate AD severity patients. In contrast, Park (2017) did not find differences in MMSE performance for post-stroke patients with cognitive deficits. However, Park and colleagues (2017) defined cognitive deficit as MMSE < 25. Of the 11 patients with cognitive deficit, 7 had MMSE scores between 18–25 and four patients had an MMSE score <18. Thus, the difference in study findings may be because the patients with cognitive deficits in Park's (2017) study were not as impaired as the "severe AD" group in Corotenuto's (2018) study. Considering these findings, the MMSE may be insufficiently valid for TNP with patients who are severely cognitively impaired or in late-stages of AD or other dementias.

Teleneuropsychology Equipment

Nearly all validity studies utilized desktop or laptop computers, which were set up in a room inside a clinic or hospital. Only two studies were conducted in patient homes (Abdolahi et al., 2018; Stillerova et al., 2016) and only one study used a smartphone (Park et al., 2017). In this study, testing was done in the clinic and the smartphone was owned by the researchers

and set up on a tripod for “hands-free” interacting. Thus, the patient did not have direct control of the smartphone during the assessment nor was there concern for distractions from notifications. Of the 11 patients who used their own equipment in Stillerova and colleagues’ (2016) study, 9 used computers and only 2 used a smartphone or tablet.

There is a clear temporal trend in which the technology used in these validity studies becomes increasingly sophisticated and convenient for patients; from the television unit used in Montani (1997) to the PC-based teleconferencing system used by Loh (2007), the tablet laptop used by Vahia (2015), the smartphone used by Park (2017), and the patient’s own in-home equipment used by Abdolahi (2018) and Stillerova (2016). Recent studies also began to use cloud-based videoconferencing. For example, Lindauer utilized Cisco’s Jabber Telepresence platform, Chapman (2019) used Zoom, and Stillerova (2016) used Skype or Google+. Generally, all studies published after 2007 had sufficiently high-speed internet connections (> 25 mbit/s).

Taken together, there is sufficient evidence for the utility of PC and laptop computers in TNP. However, there is insufficient evidence for the use of smartphones. More recent studies are starting to use cloud-based communication services, which does not seem contraindicated so long as there is a sufficiently fast and reliable internet connection.

Discussion

In response to the COVID-19 pandemic, the field of neuropsychology must rapidly evolve in response to public health concerns and social distancing directives. TNP, which was in its nascent stages at the time the outbreak began, is likely to become the preferred modality for both research and clinical practice. This systematic review provides an updated review of the available validity studies published since Brearly and colleagues’ (2017) meta-analysis. Importantly, it also offers a comprehensive outline of test-level validity data to support neuropsychologists’ informed decision-making as they select neuropsychological measures for TNP assessments with older adults.

In addition to nine articles included in Brearly and colleagues’ (2017) meta-analysis, ten additional studies were identified and included in this updated systematic review of TNP for older adults. The included studies ranged from small (N = 8) to large (N = 202) from five different countries (US, Canada, Australia, Italy, Korea), with a variety of clinical populations (MCI, dementia, PD, HD, HC, Rehabilitation, Psychiatry).

The cognitive screeners, MMSE and MoCA, had the best support; the ADAS-cog and RBANS showed promise, but had limited validity studies (n = 2 and n =1, respectively). Only one study assessed the validity of intelligence tests (Hildebrand et al., 2004). While the results of the validity testing were promising, the study sample consisted of healthy volunteers; thus, generalizability to a clinical sample should be cautioned. There appears to be moderate-to-strong evidence for all aspects of the Digit Span Task (forward, backward, total). With regards to language, the BNT-60 item and BNT-15 appeared to show good support for TNP assessments; though, Brearley and colleagues (2017) found a small, but significant difference in their quantitative review of tests (g = .10), which may be a reflection

of inconsistent performance across the alternate forms of the BNT-15 item test. Letter fluency showed excellent support for TNP validity. In contrast, due to variability across study findings, category fluency only had moderate support for validity in TNP. This may be due to the single-trial nature of the test (e.g., Animals) which may be more susceptible to performance variability in comparison to the three-trial nature of letter fluency. Thus, using a multi-trial category fluency test may be warranted. With respect to memory, the HVLT-R has strong support for validity in TNP assessments. For measurement of executive functioning, the Clock Drawing task only had moderate support due to variability in test findings and less-than-optimal validity metrics (e.g., highest ICC = .71). Unfortunately, many traditional aspects of executive functioning (e.g., set-shifting and mental flexibility, verbal inhibitory control, abstract problem-solving) were not included or well-represented in these studies and therefore have not been formally validated, to date. All other tests outlined in this review (e.g., Oral Trail Making Test A & B, BVMT-R, Token Test, Ponton-Satz Spanish Naming Test, Brief Test of Attention) only had one validity study. These tests generally showed promising results, but due to the small sample sizes of these individual studies, the evidence of validity of these measure for TNP assessments is limited. There is a lack of strong TNP validity support for measures of executive functioning and processing speed, which are often instrumental for differential diagnosis.

Although some tests show more promise over others, overall, the findings from this review suggest there is good evidence for the validity of TNP assessments for older adults. This is consistent with the findings from other meta-analyses and systematic reviews (Adams et al., 2020; Brearly et al., 2017; Costanzo et al., 2020). It is important to note that based on the available validity studies, no specific test was considered *invalid* for TNP assessments. Rather, some measures (e.g., MMSE) have much more convincing validity support than other measures with single validity studies (e.g., WAIS-III Matrix Reasoning, Hildebrand et al., 2014), smaller sample sizes (e.g., Picture Description, Vestal et al., 2006), or inconsistent findings (e.g., Clock Drawing Test). It is also important to note that there is TNP validity evidence for only a small fraction of the neuropsychological assessments that are well-validated for FTF assessments. That is not to say that TNP should eschew these other assessments. Rather, it would be prudent to create a protocol that includes validated instruments as well as supplementary measures chosen based on the specific needs of the patient and referral question (see Implementing Teleneuropsychology Into Clinical Practice).

Teleneuropsychology Logistics and Equipment

Nearly all the videoconference portions of the validity studies took place on-site of a clinic or hospital. Moving forward with in-home TNP assessments, the uncontrolled environment of the patient's home may present additional challenges. For example, the presence of unanticipated distractions and interruptions in an uncontrolled environment, the variability in quality of in-home telecommunications and computer equipment, and effects of differences in technological expertise among patients assessed in the TNP environment may all affect the validity and accuracy of test results. As one example, Stillerova and colleagues (2016) comment that a patient in their study appeared distracted by noise from their home, possibly explaining their slightly lower MoCA performance during videoconference. With these issues in mind, special care should be taken to ensure the patient is tested in a quiet

and private space where risks of distractions and interruptions can be mitigated. In-home TNP assessments also pose a unique challenge for families and caretakers who may have little familiarity with technology such as web cameras and cloud-based videoconferencing services. Thus, it will be critical to involve caretakers during the setup phase to facilitate TNP (Lindauer et al., 2017; Radhakrishnan, Xie, & Jacelon, 2016), especially as TNP appears to be less valid with advanced stages of neurological disease conditions (Carotenuto et al., 2018). Finally, the TNP model that solely relies on existing equipment in the patient's home limits the neuropsychologist's ability to fully monitor the patient's behavior due to reduced visibility from a single camera angle. With less control over, and awareness of, the physical testing environment, the neuropsychologist must rely more heavily on the patient to comply with task instructions and avoid participating in disallowed strategies such as note-taking.

Regarding equipment, nearly all studies used PC or laptops during the TNP assessments. Only two recent studies used the patient's own in-home equipment (Abdolahi et al., 2018; Stillerova et al., 2016), and one study used smartphones for video conferencing (Park et al., 2017), which was owned by the researchers and set up on a tripod. Even for the in-home studies, most of the patients used their own computers rather than smartphones/tablets (Stillerova et al., 2016). There is ample evidence for the use of PC and laptops for TNP assessments, but the use of the patient's own smartphone is not yet validated and may be contraindicated given their small size and risk for pop-up notifications (e.g., text messages) during use. Similar concerns exist for tablets.

Neuropsychologists should be aware that there are, as yet, no existing validity studies supporting the use of stand-alone computerized neuropsychological assessment devices (CNAD's) in the TNP environment. Such devices, which include turnkey assessment batteries such as ImPACT, CNS Vital Signs, CANTAB, HeadMinder, ANAM, etc., mostly require local software installation and are typically administered in a supervised office environment that enables FTF interaction with an examiner to introduce the battery and to troubleshoot difficulty. There are a number of critical issues that must be addressed before these devices can be utilized in the provision of neuropsychological services (Bauer et al., 2012), and additional unknowns exist when they are taken outside the clinic or office environment and utilized in patient homes. Until these devices can be evaluated in the sort of socially distanced TNP environment (in which the examinee is unaccompanied by the professional team), neuropsychologists should proceed with extreme caution in using these instruments.

Cultural Considerations

There was an underrepresentation of racial and ethnic minorities in most reviewed TNP validity studies. One validity study included a sample of Spanish-Speaking Hispanics, which showed promising results for TNP assessments administered in Spanish. Two studies had a large proportion of American Indians (Wadsworth et al., 2016; Wadsworth et al., 2018), which also showed promising validity results. Otherwise, ethnic minorities were largely underrepresented in these validity studies. This is problematic as older African Americans and Hispanics are less likely to use technology for health-related services compared to

Caucasians (Mitchell, 2019). Furthermore, socio-economic factors lead to barriers in access to technological utilities. For example, a 2019 Pew Research Center survey found that African American and Hispanic adults are less likely to than Caucasians to report owning a desktop or laptop computer (Perrin & Turner, 2019). Furthermore, ethnic minorities in this survey were less likely to report having access to high-speed broadband internet (Perring & Turner, 2019).

Libraries are useful in providing reliable access to computers and internet, especially for African Americans (Horrigan, 2016). However, given the closure of many non-essential buildings due to COVID-19, this may not be a viable option. Luckily, research suggests that African Americans and Hispanics reported owning smartphones at a similar rate to Caucasians (Peririn & Turner, 2019). Thus, this may be a viable means of bridging the technological divide. Although utilization of smartphones for TNP assessments are not indicated at this time, it may be a necessary option to provide underserved populations access to these services during this ongoing pandemic. While we are not making a particular endorsement of the use of smartphones to assist in TNP testing in minority populations without proper equipment or reliable internet, we implore clinicians to utilize cultural competence and make a situationally-specific decision that balances our ethical obligations of Justice and Respect for People's Rights and Dignity to provide equal access to such services, with our obligation of Beneficence and Nonmaleficence (American Psychological Association, 2017) as testing via unvalidated media may lead to greater likelihood of harm (e.g., higher likelihood of false positive diagnoses, stricter functional recommendations and restrictions). In the absence of reliable internet, computers, or smartphones, a neuropsychological screening via telephone may also be indicated (e.g., see Lachman, Agrigoroaei, Tun, & Weaver, 2014).

Strengths and Limitations

A limitation of the current review is the relatively narrow scope of studies reviewed. First, only counter-balanced cross-over designs were included in this review (with two exceptions) because cross-over designs are optimal for the assessment of reliability compared to cross-sectional or single-arm randomized design studies. Other studies have used extensive TNP batteries and have reported good levels of diagnostic utility and patient satisfaction (e.g., Barton et al., 2011; Harrell et al., 2014). However, since these studies were not cross-over designs, the reliability of these assessments administered via TNP could not be compared to FTF performance.

Secondly, Medicare (at the time of manuscript preparation) requires both audio and visual input for a healthcare delivery episode to be considered telehealth. Telephone screening and assessments would not meet criteria for telehealth and would not be reimbursed as such. Therefore, a full analysis of validated telephone screeners and assessments was not attempted as part of this review. If needed, there are several, large studies with well-validated telephone assessments and screeners. For example, Lachman et al. (2014), conducted a large (N = 4,268) study of community dwelling adults aged 32–84 that showed good convergent validity with gold-standard cognitive tests that were assessed FTF. Their battery included frequently used assessments such as RAVLT, Digit Span, and the Stop

Go and Switch Task (e.g., see Bunker et al., 2017; Mitsis et al., 2010; Wynn, Sha, Lamb, Carpenter, & Yochim, 2019 for additional telephone validity studies).

Finally, it was beyond the scope of this review to assess the relative validity of web-based neuropsychological assessments that were not conducted via cross-over design. For example, the NeuroCognitive Performance Test is an exclusively web-based assessment with a normative data sample of 130,140 healthy volunteers who took the assessments remotely and without supervision (Morrison, Simone, Ng, & Hardy, 2015). In this study, a subset of 1,493 individuals with self-reported MCI or AD performed significantly worse on the overall score than healthy age-, gender-, and education-matched controls. Additionally, Millisecond Test Library has a 651 test paradigm that can be administered via a web-based portal and all tests are free with a paid license with Inquisit (https://www.millisecond.com/download/library/). Future research should determine if traditional neuropsychological assessments delivered over videoconference is better at diagnosing various neurocognitive disorders than these web-based assessment platforms.

Nonetheless, this review builds upon the work completed by Brearly and colleagues (2017) and provides an update of validity studies conducted in the past four years. Notably, this is the first to provide a qualitative review of validity studies, stratified by individual tests. This will allow clinicians and researchers to make more informed decisions when selecting their test battery for TNP assessments. This is also the first review to outline equipment used in TNP validity studies. This is of particular importance given (1) the proliferation of in-home smart devices that may facilitate telehealth services; and (2) under the current pandemic, in-home assessments are now reimbursed by Medicare.

Implementing Teleneuropsychology Into Clinical Practice

The University of Florida has created a provisional protocol for the assessment of patients via TNP after consultation with multiple board-certified neuropsychologists in clinical practice, a review of best-practice guidelines (Maria C. Grosch, Gottlieb, & Cullum, 2011), and after considering the advice of Dr Munro Cullum, an expert in the field of TNP (Cullum, Bellone, & Van Patten, 2020), The protocol calls for a two-stage, tiered process of triaging patients and determining their suitability for TNP assessments. As suggested by Cullum (Cullum, Bellone, & Van Patten, 2020), patients will first be called to obtain a brief clinical history and to determine the appropriateness of a TNP evaluation based on the referral question and presenting problems. For example, patients who are hearing impaired, require motor testing (e.g., stroke, epilepsy patients), or where visual memory/visuomotor is important to answer the referral question (e.g., posterior cortical atrophy, epilepsy), may be deferred for in-person assessment at a later time. The preliminary screening will also serve to determine if the patient has access to equipment suitable for TNP assessment (e.g., reliable high-speed internet; PC or laptop with webcam). A brief cognitive screener (Modified - Telephone Interview for Cognitive Status; TICS-M) will be administered over the phone to provide an estimate of current cognitive functioning and collateral informant will be asked to complete the Functional Assessment Questionnaire (FAQ) to assess for impairments in activities of daily living.

Patients who score in the normal range on the TICS-M and FAQ will be triaged as low priority and deferred for FTF or TNP assessments at a later time. Patients who perform below clinical cut-offs will be assessed for suitability of TNP assessments. If reliable equipment and suitable testing environment are available, the patient will be scheduled for a TNP assessment. Assessments via smartphones will not be allowed unless it is the only means to provide a necessary assessment that cannot otherwise be completed via telephone.

TNP assessments will be conducted on the cloud-based platform, Zoom-PHI, which is secure and approved for sharing private health information. Zoom also allows multiple people to participate in a videoconference session with or without video input. Therefore, trainees can unobtrusively observe assessments and meaningfully contribute to the evaluation (e.g., live scoring of assessments, report writing). Zoom also has a screen-share feature, which allows the patient to see stimuli saved on the examiner's computer. In accordance with publisher permissions, select visually-based stimuli will be converted to a power point presentation and shown to the patient via the screen-share feature.

After beginning the telehealth appointment, the patient will be forewarned about potential technological mishaps, such as loss of connection or audio glitches. They will be instructed to re-initiate the Zoom session should they lose connection mid-appointment, and will be asked to provide a phone number so they can be reached if technical issues persist. Before beginning the assessment, the patient will be asked to find a private, quiet, distraction-free setting for the assessment and to remove all writing utensils from their area if drawing tasks are not included in the protocol. They will be encouraged to wear sensory aids during the appointment if needed (e.g., glasses, hearing aids). The session will begin with a quick tutorial to optimize video and sound quality (e.g., screen resolution and volume). Through screen-share, the patient will be shown a single image and asked if it is clear, blurry, or distorted in any manner. If the image appears blurry or distorted to the patient despite adequate vision otherwise, trouble-shooting strategies to improve internet bandwidth will be suggested (e.g., closing all other programs on their computer or tablet, moving closer to the router in their home). If none of these strategies are successful, testing may need to be modified to remove all tasks with visual-stimuli. With respect to volume optimization, the patient will be read a short passage and will be instructed to adjust the volume on their device to a comfortable level during that time. After they have optimized the volume level on their device, a brief auditory discrimination screening will be conducted to evaluate the patient's ability to discriminate between phonetically-similar words. The hearing screener was informally developed by a Speech-Language Pathologist (L. Altmann, personal communication, March 25, 2020) who used a confusion matrix to find word pairs that would be phonetically difficult to distinguish (e.g., push vs. bush; team vs. deem) (http://people.cs.uchicago.edu/~dinoj/research/wangbilger.html). The word pairs will simultaneously be presented on the screen and the patient will be asked to identify the word that was spoken by the examiner. The patient's performance on this task will be recorded for consideration during clinical interpretation of assessment results.

In regard to the evaluation, a core TNP battery was developed based on the validity studies. That is, most TNP batteries will include the HVLT-R, Digit Span, BNT, Letter Fluency, and Category Fluency. Providers will then decide on additional tests to administer based on

clinical preference and necessity given the referral question. To reduce potential distractions, especially from stimulus-bound patients, the patients will be asked to fold and place on the ground any visual stimuli they complete (e.g., Clock Drawing Test). To reduce privacy concerns for examiners who conduct assessments off-site (e.g., their own home), all scoring and verbatim responses will be completed on electronically modified score sheets that will be stored in a private server maintained by the University.

Conclusions

In response to the COVID-19 pandemic, neuropsychological assessments via videoconference appear to be a valid means of assessing cognitive functioning in older adults. This review identified several measures, assessing various cognitive domains, that have been validated for TNP based on counter-balanced cross-over designs in multi-ethnic and diagnostically diverse samples. Challenges to implementing TNP assessments in the current climate include an unfamiliarity with technological instruments, lack of access to reliable equipment, and reliance on caretakers (especially for severe cases). Due to social distancing guidelines, the current healthcare environment discourages any face-to-face interaction between patients and practitioners, thus making the situation slightly different from the typically-studied arrangement in which a teleneuropsychology visit is proctored remotely by a member of the healthcare staff. While we are currently in unchartered waters, this represents an opportunity for neuropsychologists to develop ways to evaluate the efficacy of this approach while providing valuable services to our patients and colleagues.

Acknowledgements

A special thank you to Andrea Mejia, Nicole Evangelista, and the rest of the faculty members in the department of Clinical and Health Psychology at the University of Florida for their assistance with this project.

References

*Abdolahi AB, M. T.;Darwin KC; Venkataraman V; Grana MJ; Dorsey ER; Biglan KM (2016). A feasibility study of conducting the Montreal Cognitive Assessment remotely in individuals with movement disorders. Health Informatics J, 22(2), 304–311. doi:10.1177/1460458214556373 [PubMed: 25391849]

Adhanom Ghebreyesus T (2020). WHO Director-General's opening remarks at the media briefing on COVID-19 – 20 March 2020 [Transcript]. Retrieved from https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-atthe-media-briefing-on-covid-19---11-march-2020

Adams JL, Myers TL, Waddell EM, Spear KL, & Schneider RB (2020). Telemedicine: a Valuable Tool in Neurodegenerative Diseases. Current Geriatrics Reports doi:10.1007/s13670-020-00311-z

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct (2002, amended effective June 1, 2010, and January 1, 2017) https://www.apa.org/ethics/code/

Barton C, Morris R, Rothlind J, & Yaffe K (2011). Video-telemedicine in a memory disorders clinic: evaluation and management of rural elders with cognitive impairment. Telemed J E Health, 17(10), 789–793. doi:10.1089/tmj.2011.0083 [PubMed: 22023458]

Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, & Naugle RI (2012). Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. Arch Clin Neuropsychol, 27(3), 362–373. doi:10.1093/arclin/acs027 [PubMed: 22382386]

Bland JM, & Altman DG (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1(8476), 307–310. [PubMed: 2868172]

Brearly TW, Shura RD, Martindale SL, Lazowski RA, Luxton DD, Shenal BV, & Rowland JA (2017). Neuropsychological Test Administration by Videoconference: A Systematic Review and Meta-Analysis. Neuropsychol Rev, 27(2), 174–186. doi:10.1007/s11065-017-9349-1 [PubMed: 28623461]

Bunker L, Hshieh TT, Wong B, Schmitt EM, Travison T, Yee J, . . . Inouye SK (2017). The SAGES telephone neuropsychological battery: correlation with in-person measures. Int J Geriatr Psychiatry, 32(9), 991–999. doi:10.1002/gps.4558 [PubMed: 27507320]

*Carotenuto A, Rea R, Traini E, Ricci G, Fasanaro AM, & Amenta F (2018). Cognitive Assessment of Patients With Alzheimer's Disease by Telemedicine: Pilot Study. JMIR Ment Health, 5(2), e31. doi:10.2196/mental.8097 [PubMed: 29752254]

Centers for Disease Control and Prevention. (2020). Coronavirus disease 2019 (COVID-19): Older adults Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/specificgroups/high-risk-complications/older-adults.html

*Chapman JE, Cadilhac DA, Gardner B, Ponsford J, Bhalla R, & Stolwyk RJ (2019). Comparing face-to-face and videoconference completion of the Montreal Cognitive Assessment (MoCA) in community-based survivors of stroke. J Telemed Telecare, 1357633x19890788. doi:10.1177/1357633x19890788

Cicchetti DV (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6(4), 284–290. doi:10.1037/1040-3590.6.4.284

Ciemins EL, Holloway B, Coon PJ, McClosky-Armstrong T, & Min SJ (2009). Telemedicine and the mini-mental state examination: assessment from a distance. Telemed J E Health, 15(5), 476–478. doi:10.1089/tmj.2008.0144 [PubMed: 19548827]

Cohen J (1988). Statistical power analysis for the behavioral sciences. Abingdon England: Routledge.

Costanzo MC, Arcidiacono C, Rodolico A, Panebianco M, Aguglia E, & Signorelli MS (2020). Diagnostic and interventional implications of telemedicine in Alzheimer's disease and mild cognitive impairment: A literature review. Int J Geriatr Psychiatry, 35(1), 12–28. doi:10.1002/gps.5219 [PubMed: 31617247]

Coronavirus Preparedness and Response Supplemental Appropriations Act, 2020, H.R. 6074, 116th Cong., 2nd Sess. (2020).

*Cullum M (Guest), Bellone J & Van Patten R (Producers) (2020, March 25). Teleneuropsychology – With Dr. Munro Cullum [Audio Podcast]. Retrieved from: https://www.navneuro.com/41-teleneuropsychology-with-dr-munro-cullum/

*Cullum CM, Weiner MF, Gehrmann HR, & Hynan LS (2006). Feasibility of telecognitive assessment in dementia. Assessment, 13(4), 385–390. doi:10.1177/1073191106289065 [PubMed: 17050908]

Cullum MC, Hynan LS, Grosch M, Parikh M, & Weiner MF (2014). Teleneuropsychology: evidence for video teleconference-based neuropsychological assessment. J Int Neuropsychol Soc, 20(10), 1028–1033. doi:10.1017/s1355617714000873 [PubMed: 25343269]

DeYoung N, & Shenal BV (2019). The reliability of the Montreal Cognitive Assessment using telehealth in a rural setting with veterans. J Telemed Telecare, 25(4), 197–203. doi:10.1177/1357633x17752030 [PubMed: 29320916]

Ding H, Hu GL, Zheng XY, Chen Q, Threapleton DE, & Zhou ZH (2015). The method quality of cross-over studies involved in Cochrane Systematic Reviews. PLoS One, 10(4), e0120519. doi:10.1371/journal.pone.0120519 [PubMed: 25867772]

*Galusha-Glasscock JM, Horton DK, Weiner MF, & Cullum CM (2015). Video Teleconference Administration of the Repeatable Battery for the Assessment of Neuropsychological Status. Archives of Clinical Neuropsychology, 31(1), 8–11. doi:10.1093/arclin/acv058 [PubMed: 26446834]

*Grosch MC, Gottlieb MC, & Cullum CM (2011). Initial Practice Recommendations for Teleneuropsychology. Clin Neuropsychol, 25(7), 1119–1133. doi:10.1080/13854046.2011.609840 [PubMed: 21951075]

Grosch MC, Weiner MF, Hynan LS, Shore J, & Cullum CM (2015). Video teleconference-based neurocognitive screening in geropsychiatry. Psychiatry Res, 225(3), 734–735. doi:10.1016/j.psychres.2014.12.040 [PubMed: 25596957]

Harrell KM, Wilkins SS, Connor MK, & Chodosh J (2014). Telemedicine and the evaluation of cognitive impairment: the additive value of neuropsychological assessment. J Am Med Dir Assoc, 15(8), 600–606. doi:10.1016/j.jamda.2014.04.015 [PubMed: 24913209]

*Hildebrand R, Chow H, Williams C, Nelson M, & Wass P (2004). Feasibility of neuropsychological testing of older adults via videoconference: implications for assessing the capacity for independent living. J Telemed Telecare, 10(3), 130–134. doi:10.1258/135763304323070751 [PubMed: 15165437]

Horrigan JB (2016). Library usage and engagement. Libraries 2016 Retrieved from https://www.pewresearch.org/internet/2016/09/09/library-usage-and-engagement/

Koo TK, & Li MY (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of chiropractic medicine, 15(2), 155–163. doi:10.1016/j.jcm.2016.02.012 [PubMed: 27330520]

Lachman ME, Agrigoroaei S, Tun PA, & Weaver SL (2014). Monitoring cognitive functioning: psychometric properties of the brief test of adult cognition by telephone. Assessment, 21(4), 404–417. doi:10.1177/1073191113508807 [PubMed: 24322011]

*Lindauer A, Seelye A, Lyons B, Dodge HH, Mattek N, Mincks K, . . . Erten-Lyons D (2017). Dementia Care Comes Home: Patient and Caregiver Assessment via Telemedicine. Gerontologist, 57(5), e85–e93. doi:10.1093/geront/gnw206

*Loh PK, Donaldson M, Flicker L, Maher S, & Goldswain P (2007). Development of a telemedicine protocol for the diagnosis of Alzheimer's disease. J Telemed Telecare, 13(2), 90–94. doi:10.1258/135763307780096159 [PubMed: 17359573]

*Loh PK, Ramesh P, Maher S, Saligari J, Flicker L, & Goldswain P (2004). Can patients with dementia be assessed at a distance? The use of Telehealth and standardised assessments. Intern Med J, 34(5), 239–242. doi:10.1111/j.1444-0903.2004.00531.x [PubMed: 15151669]

McEachern W, Kirk A, Morgan DG, Crossley M, & Henry C (2008). Reliability of the MMSE administered in-person and by telehealth. Can J Neurol Sci, 35(5), 643–646. doi:10.1017/s0317167100009458 [PubMed: 19235450]

Menon AS, Kondapavalru P, Krishna P, Chrismer JB, Raskin A, Hebel JR, & Ruskin PE (2001). Evaluation of a portable low cost videophone system in the assessment of depressive symptoms and cognitive function in elderly medically ill veterans. J Nerv Ment Dis, 189(6), 399–401. doi:10.1097/00005053-200106000-00009 [PubMed: 11434642]

Mitsis EM, Jacobs D, Luo X, Andrews H, Andrews K, & Sano M (2010). Evaluating cognition in an elderly cohort via telephone assessment. Int J Geriatr Psychiatry, 25(5), 531–539. doi:10.1002/gps.2373 [PubMed: 19697298]

Moher D, Liberati A, Tetzlaff J, & Altman DG (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med, 6(7), e1000097. doi:10.1371/journal.pmed.1000097 [PubMed: 19621072]

*Montani C, Billaud N, Tyrrell J, Fluchaire I, Malterre C, Lauvernay N, . . . Franco A (1997). Psychological impact of a remote psychometric consultation with hospitalized elderly people. J Telemed Telecare, 3(3), 140–145. doi:10.1258/1357633971931048 [PubMed: 9489108]

Morrison GE, Simone CM, Ng NF, & Hardy JL (2015). Reliability and validity of the NeuroCognitive Performance Test, a web-based neuropsychological assessment. Front Psychol, 6, 1652. doi:10.3389/fpsyg.2015.01652 [PubMed: 26579035]

Office for Civil Rights. (2020). Notification of enforcement discretion for telehealth remote communications during the COVID-19 nationwide public health emergency Retrieved from https://www.hhs.gov/hipaa/for-professionals/special-topics/emergencypreparedness/notification-enforcement-discretion-telehealth/index.html#

Parikh M, Grosch MC, Graham LL, Hynan LS, Weiner M, Shore JH, & Cullum CM (2013). Consumer acceptability of brief videoconference-based neuropsychological assessment in older individuals with and without cognitive impairment. Clin Neuropsychol, 27(5), 808–817. doi:10.1080/13854046.2013.791723 [PubMed: 23607729]

*Park HY, Jeon SS, Lee JY, Cho AR, & Park JH (2017). Korean Version of the Mini-Mental State Examination Using Smartphone: A Validation Study. Telemed J E Health, 23(10), 815–821. doi:10.1089/tmj.2016.0281 [PubMed: 28422578]

Perrin A & Turner E (2019). Smartphones help blacks, Hispanics bridge some – but not all digital gaps with whites Retrieved from https://www.pewresearch.org/facttank/2019/08/20/smartphones-help-blacks-hispanics-bridge-some-but-not-all-digitalgaps-with-whites/

Radhakrishnan K, Xie B, & Jacelon CS (2016). Unsustainable Home Telehealth: A Texas Qualitative Study. Gerontologist, 56(5), 830–840. doi:10.1093/geront/gnv050 [PubMed: 26035878]

Smith A (2015). Chapter one: A portrait of smartphone ownership. U.S. Smartphone Use in 2015 Retrieved from https://www.pewresearch.org/internet/2015/04/01/chapter-one-aportrait-of-smartphone-ownership/#cancel-phone

*Stillerova T, Liddle J, Gustafsson L, Lamont R, & Silburn P (2016). Could everyday technology improve access to assessments? A pilot study on the feasibility of screening cognition in people with Parkinson's disease using the Montreal Cognitive Assessment via Internet videoconferencing. Aust Occup Ther J, 63(6), 373–380. doi:10.1111/1440-1630.12288 [PubMed: 27059159]

Turkstra LS, Quinn-Padron M, Johnson JE, Workinger MS, & Antoniotti N (2012). In-person versus telehealth assessment of discourse ability in adults with traumatic brain injury. J Head Trauma Rehabil, 27(6), 424–432. doi:10.1097/HTR.0b013e31823346fc [PubMed: 22190010]

Turner TH, Horner MD, Vankirk KK, Myrick H, & Tuerk PW (2012). A pilot trial of neuropsychological evaluations conducted via telemedicine in the Veterans Health Administration. Telemed J E Health, 18(9), 662–667. doi:10.1089/tmj.2011.0272 [PubMed: 23050802]

*Vahia IV, Ng B, Camacho A, Cardenas V, Cherner M, Depp CA, . . . Agha Z (2015). Telepsychiatry for Neurocognitive Testing in Older Rural Latino Adults. Am J Geriatr Psychiatry, 23(7), 666–670. doi:10.1016/j.jagp.2014.08.006 [PubMed: 25708655]

*Vestal L, Smith-Olinde L, Hicks G, Hutton T, & Hart J Jr. (2006). Efficacy of language assessment in Alzheimer's disease: comparing in-person examination and telemedicine. Clin Interv Aging, 1(4), 467–471. doi:10.2147/ciia.2006.1.4.467 [PubMed: 18046923]

*Wadsworth HE, Dhima K, Womack KB, Hart J Jr., Weiner MF, Hynan LS, & Cullum CM (2018). Validity of Teleneuropsychological Assessment in Older Patients with Cognitive Disorders. Arch Clin Neuropsychol, 33(8), 1040–1045. doi:10.1093/arclin/acx140 [PubMed: 29329363]

*Wadsworth HE, Galusha-Glasscock JM, Womack KB, Quiceno M, Weiner MF, Hynan LS, . . . Cullum CM (2016). Remote Neuropsychological Assessment in Rural American Indians with and without Cognitive Impairment. Arch Clin Neuropsychol, 31(5), 420–425. doi:10.1093/arclin/acw030 [PubMed: 27246957]

Wallace BC, Small K, Brodley CE, Lau J, & Trikalinos TA (2012). Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. Paper presented at the Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, Miami, Florida, USA. 10.1145/2110363.2110464

World Health Organization. (2020). Coronavirus disease 2019 (COVID-19) Situation Report 69 Retrieved from https://www.who.int/docs/default-source/coronaviruse/situationreports/20200329-sitrep-69-covid-19.pdf?sfvrsn=8d6620fa_4.

Wynn MJ, Sha AZ, Lamb K, Carpenter BD, & Yochim BP (2019). Performance on the Verbal Naming Test among healthy, community-dwelling older adults. Clin Neuropsychol, 1–13. doi:10.1080/13854046.2019.1683232

*Yoshida K, Yamaoka Y, Eguchi Y, Sato D, Iiboshi K, Kishimoto M, . . . Kishimoto T (2019). Remote neuropsychological assessment of elderly Japanese population using the Alzheimer's Disease Assessment Scale: A validation study. J Telemed Telecare, 1357633x19845278. doi:10.1177/1357633x19845278

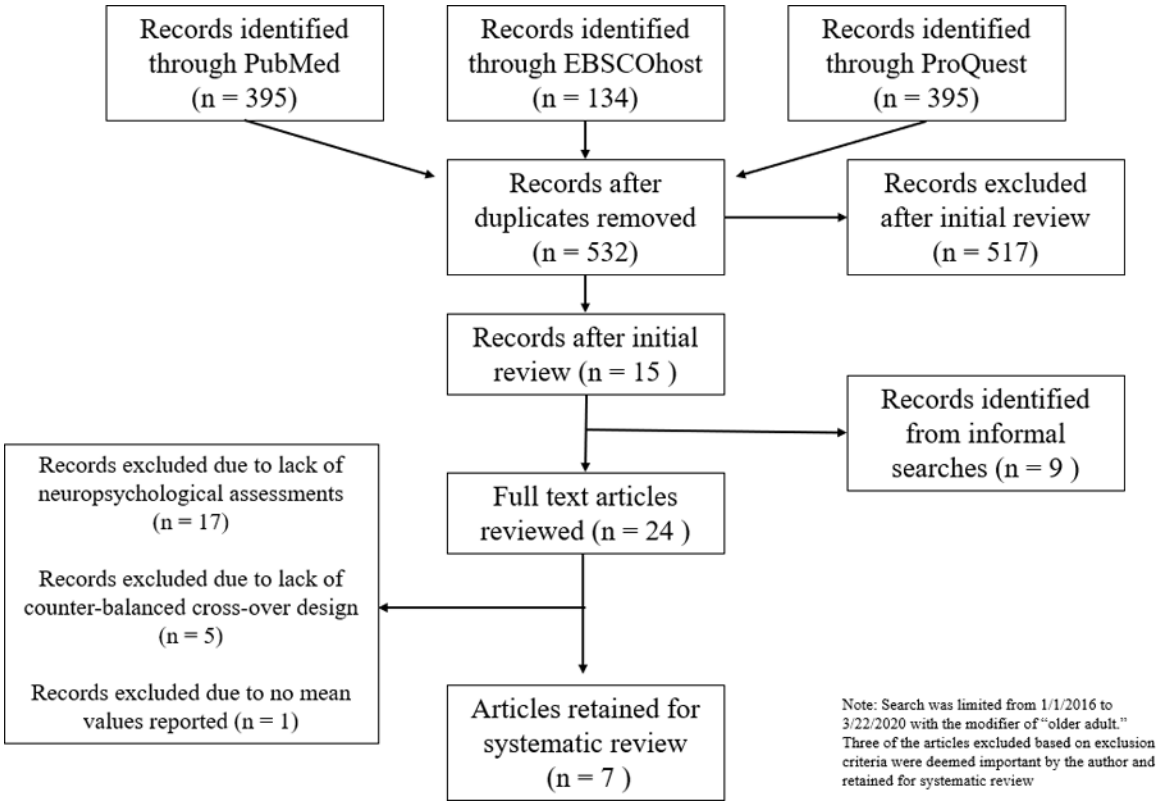

Figure 1.
Flowchart of Study Selection

Table 1.

Results of Brearly et al., (2017) Meta-Analysis

Test	k	N	Hedges *g*	*Q*	I^2 (%)
BNT or BNT-15	4	329	−0.12***	1.76	0
Semantic Fluency	3	319	−0.08	5.77	65.34
Clock Drawing	5	335	−0.13	12.6	68.25
Digit Span	5	359	−0.05	9.38	57.34
List Learning (total)	3	313	0.1	4.59	56.46
MMSE	7	380	−0.4	3.36***	80.24
Letter Fluency	5	356	−0.02	1.41	0
Synchronous Dependent Tests	NR	NR	NR[a]	56.42***	82.28
Non-Synchronous Dependent Test	NR	NR	−0.10***	12.99	38.43
Overall Results	**12**	**497**	**−0.03**	**55.67*** **	**80.24**

Note. Adopted from Brearly et al. (2017) published in *Neuropsychology Review*

BNT = Boston Naming Test; MMSE = Mini Mental State Examination; NR = Not Reported Synchronous refers to timed tests or single-trial tests where repetition could confound results (e.g., digit span).

* = $p < .05$,

** = $p < .01$,

*** = $p < .001$

[a] Authors did not provide an effect size estimate due to significant between-study heterogeneity

Table 2.

Brief Tests of Global Cognitive Functioning

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
MoCA	Lindauer (2017)	N = 66, n = 33 AD, n =33 caretakers; Mage $_{AD}$ = 71.6 (SD = 11.6); 61% Female; "mostly" Caucasian	Using their computer and a camera via Cisco's Jabber TelePresence platform; iPads loaned if patients did not have a computer	-	2 weeks	ICC = 0.93 (excellent)	Visuospatial materials enlarged and mailed to patient; patient held up stimuli to camera for scoring
	Abdolahi (2018)[a]	N = 17 PD/HD patients; Mage = 61.18, 70.61% Female	Patient's own in-home equipment	High Speed	7 months for PD patients; 3 months for HD patients	Slight, but non-significant increase in scores from FTF to VC, moreso in PD group. ICC = .59 (good), Cronbach's alpha = .74 (adequate), Pearson's r = .59 (strong)	Visuosptial and naming sub-section emailed to patient
	Chapman (2019)	N = 48 stroke survivors from Australia; Mage = 64.6 (SD = 10.1), Meducation = 13.7 (SD = 3.3); 46% Female	Two laptops, provided by researchers, located in separate rooms at the same location; ideoconference sessions were conducted using the cloud-based videoconferencing Zoom	384 kbit/s	2 weeks	No difference in total scores across testing modalities (t(47) = .44, p - .658, d = 0.06); similar findings across MoCA domain scores. ICC = .615 (good); 72.9% of participants classified consistently across conditions (normal vs. impaired). Neither age, computer literacy, or hospital depression/anxiety predicted difference scores between conditions	A MoCA response form including only the visuospatial/executive and naming items was in an envelope at the participant's location.
	Stillerova (2016) [a]	N = 11 Austrailian patients with PD; Median age = 69; 36% Female	Patient's own devices using Skype or Google+ Hangouts; Completed at their own home	-	7 days	Median difference in scores was 2.0 points out of 30 (IQR = 1.0–2.5); 2 patients changed from 'impaired' to 'normal' from FTF to videoconference, 1 patients went from 'normal' to 'impaired'	Non-randomized design (FTF then Videoconference); Patients given a sealed envolope with the visual stimuli;
MMSE	Cullum (2014) *	N = 202, n = 83 MCI/AD, n = 119 healthy controls; Mage = 68.5 (SD = 9.5), Meducation = 14.1 (SD = 2.7); 63% Female	H.323 PC - Based Videoconferencing System (Polycom™ iPower 680 Series) that was set up in two non-adjacent rooms	High Speed	Same Day	Similar mean scores between two testing modalities. ICC = .798 (excellent)	The visuospatial measures were scored via the television monitor by asking participants to hold up their paper in front of camera
	Cullum (2006) *	N = 33, n = MCI, n = AD; Mage = 73.5 (SD = 6.9), Meducation = 15.1 (SD = 2.7); 33% Female; 97% Caucasian	An H.323 PC-based Videoconferencing System was set up in two nonadjacent rooms using a Polycom iPower 680 Series videoconferencing system (2 units)	High Speed	Same Day	Similar mean scores between two testing modalities. ICC = .88 (excellent)	The visuospatial measures were scored via the television monitor by asking participants to hold up their paper in front of camera

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
	Loh (2004) *	N = 20 Mixed Clinical Sample from Australia; Mage = 82; 80% Female	VCON cruiser, version 4.0, videoconferencing unit with Sony D31 PTZ camera; used onsite	128–384 kb/s	-	Similar mean scores between two testing modalities. Correlation = .90. Limits of agreement = −4.6 – 4.0; Reliability increased when 4 patients with delerium were removed from analyses	
	Montani (1997) *	N = 14 Mixed rehabilitation; Mage = 88 (SD = 5), 46.7% Female	Camera, television screen, and microphone in an adjacent room	Coaxial	8 days	Small, but significant difference between testing modalities with better FTF performance (p = .003); strong correlation of scores between testing modalities (r = .95)	
	Loh (2007) *	N = 20 Memory Disorder Sample from Australia; Mage = 79; 55% Female	PC-based videoconferencing equipment (Cruiser, version 4, VCON). Conducted on-site	384 kbit/s	-	Similar mean scores between two testing modalities. ICC = .89 (excellent). Limits of agreement = −1.89 – 0.04). Kappa of physicians making AD diagnosis face-to-face and in person = .80 ($p < 0.0001$)	
	Carotenuto (2018)	N = 28 AD patients from Italian Memory Disorders Clinic; Mage = 75.39; Meducation = 7.61 (SD = 4.07); 71% Female	Sony VAIO laptops contained an IntelCore Duo CPU P8400 2.26 GHz processor, 4 GB memory, Intel Media Accelerator X3100 graphics card, and a 17.3″ LCD LED (1920×1080) integrated screen. Completed at the hospital	100 Mbit/s	Video and FTF done at baseline, 6, 12, 18, and 24 months. Video and FTF assessments done 2 weeks apart	No mean differences in performance across testing modalities at any timepoint ($p > .05$); Video performance lower (worse) than FTF at baseline and 24 months for severe patients (MMSE 15–17), but not moderate (MMSE = 18–20) or slight (MMSE = 21–24) AD patients	
	Park (2017)	N = 30 Korean patients with stroke, Mage = 68.83 (SD = 12.95), Meducation = 11.70 (SD = 5.38); 66.6% Female	martphone (iPhone 5S) with communication via FaceTime; assessment conducted in adjacent room	100 Mbit/s	3 days	No significant differences in scores between testing modalities for Total Score (Z = −1.574, $p = .116$) or across subdomains (all $p > .05$). Spearman correlation significant (.949, $p < .001$). No significant differences between patitents with mild aphasia or dysarhtira (n = 11, p = *.039*) or patients with cognitive defict (MMSE < 25; n = 4, p = 0.104)	
	Vahia (2015) *a*	N = 22 spanish-speaking Hispanics living in US who were referred for cognitive testing by psychiatrist; Mage = 70.75, Meducation = 5.54 ; 22.75% Female; 0% Caucasian; 100% Hispanic	CODEC (coder-decoder) capable of simultaneously streaming video and content (i.e., laptop screen) on side by side monitors, remotely controlled Pan Tilt and Zoom cameras, a tablet PC laptop, videoconference microphone; and dual 26 inch LCD TVs	512 kbit/s	2 weeks	Using mixed-effects models, no significant difference between testing modalities on overall Cognitive Composite score derived from entire test battery ($F(1,37) = .31$, $p = .0579$), with slightly better (but not significantly better) performance at second testing session, irrespective of modality. No	Administered in Spanish

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
						significant difference in MMSE scores (differences in test scores across testing modalities not provided)	
	Wadsworth (2016) *	N = 84, n = 29 MCI/ dementia, n = 55 HC; Mage = 64.89 (SD = 9.73), Meducation = 12.58 (SD = 2.35); 63% Female; 0% Caucasian; 100% American Indian	Polycom iPower 680 series videoconferencing system in same facility	High Speed	Same Day	No differences in mean performance. ICC = .92 (excellent)	ubjects were asked to holdup their drawing to the camera in order for the examiner to score it immediately
	Grosch (2015) *	N = 8 geropsychiatric VA patients; 12.5% Female	An H.323 PC-based Videoconferencing System at VAMC	384 kb/s	Same Day	No difference in mean test score across conditions. ICC was non-significant and just-above “poor” (ICC =.42)	After completing the MMSE pentagons and Clock Drawings, subjects were asked to hold their paper in front of the camera so as to allow scoring by the examiner via television monitor.
ADAS-cog	Carotenuto (2018)	N = 28 AD patients from Italian Memory Disorders Clinic; Mage = 75.39; Meducation = 7.61 (SD = 4.07); 71% Female	Sony VAIO laptops contained an IntelCore Duo CPU P8400 2.26 GHz processor, 4 GB memory, Intel Media Accelerator X3100 graphics card, and a 17.3″ LCD LED (1920×1080) integrated screen. Completed at the hospital	100 Mbit/s	Video and FTF done at baseline, 6, 12, 18, and 24 months. Video and FTF assessments done 2 weeks apart	No mean differences in performance across testing modalities at any timepoint ($p > .05$); Video performance worse (higher) than FTF at all timepoints for severe patients (MMSE 15–17), but not moderate (MMSE = 18–20) or slight (MMSE = 21–24) AD patients	
	Yoshida (2019)	N = 73, n = 34 MCI/AD, n = 48 HC; Mage = 76.3 (SD = 7.6), Meducation = 13; 50.7% Female; 100% Japanese sample	Cisco TelePresenceVR System EX60, DX80, SX20 and Roomkit in an adjacent examiner room	High Speed	2 weeks - 3 months	ICC = .86 for whole sample (excellent); ICCMCI = .63 (good); ICCdementia = .80 (excellent); ICCHC = .74 (good)	
RBANS	Galusha-Glassock (2016) *	N = 18, n = 11 MCI/AD, n = 7 HC; Mage = 69.67 (SD = 7.76), Meducation = 14.28 (SD = 2.76), *MMMSE* = 26.72 (SD = 2.89); 38.8% Female; 78% Caucasian	Polycom iPower 680 series videoconferencing system in two nonadjacent rooms in the same facility	High Speed	Same Day	Mean scores for RBANS Total and all index scores were statistically similar across testing modalities. ICC for Total Score = .88 (excellent). ICC for index scores was fair visuosptial/ constructional (ICC = .59). ICC for all other index scores were excellent (ICC range .75-.90)	examiner held up the stimulus in front ofthe cameraforRBANSFigureCopy, Line Orientation, Picture Naming,and Coding.Blank paperand a penwere available in the testingroomfor the participant as wasacopy ofthe Coding sheet from the test protocol.

Notes.

* indicates a study from Brearly et al., (2017)

[a] = not a counter-balanced cross-over design (FTF then videoconference administration)

Table 3.

Test of Intelligence

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
Matrix Reasoning	Hildebrand (2004) *	N = 29 HC from Canada, Mage = 68 (SD = 8), Meducation = 13 (SD = 3), *MMMSE* = 28.9 (SD = 1.3); 73% Female	Videoconferencing systems (LC5000, VTEL) with two 81cm monitors and far-end camera control. Conducted on-site	336 kbit/s	2–4 weeks apart	No difference in mean score across conditions (Limits of agreeement −4.56 – 6.08)	Pictures of the visual test material were presented using a document camera (Elmo Visual Presenter, EV400 AF).
Vocabulary	Hildebrand (2004) *	N = 29 HC from Canada, Mage = 68 (SD = 8), Meducation = 13 (SD = 3), *MMMSE* = 28.9 (SD = 1.3); 73% Female	Videoconferencing systems (LC5000, VTEL) with two 81cm monitors and far-end camera control. Conducted on-site	336 kbit/s	2–4 weeks apart	No difference in mean score across conditions (Limits of agreeement −3.07 – 3.13)	

Notes.

* indicates a study from Brearly et al., (2017)

Table 4.

Tests of Attention/Working Memory

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
Digit Span Forward	Cullum (2014) *	N = 202, n = 83 MCI/AD, n = 119 healthy controls; Mage = 68.5 (SD = 9.5), Meducation = 14.1 (SD = 2.7); 63% Female	H.323 PC - Based Videoconferencing System (Polycom™ iPower 680 Series) that was set up in two non-adjacent rooms	High Speed	Same Day	Similar mean performance. ICC = .590 (fair)	
	Wadsworth (2016) *	N = 84, n = 29 MCI/dementia, n = 55 HC; Mage = 64.89 (SD = 9.73), Meducation = 12.58 (SD = 2.35); 63% Female; 0% Caucasian; 100% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	Small but significant difference with better face-to-face performance (t - 2.98, p = *.004).* ICC = .75 (good)	
	Vahia (2015)	N = 22 spanish-speaking Hispanics living in US who were referred for cognitive testing by psychiatrist; Mage = 70.75, Meducation = 5.54 ; 22.75% Female; 0% Caucasian; 100% Hispanic	CODEC (coder-decoder) capable of simultaneously streaming video and content (i.e., laptop screen) on side by side monitors, remotely controlled Pan Tilt and Zoom cameras, a tablet PC laptop, videoconference microphone; and dual 26 inch LCD TVs	512 kbit/s	2 weeks	Using mixed-effects models, no significant difference between testing modalities on overall Cognitive Composite score derived from entire test battery (F(1,37) = .31, p =.0579), with slightly better (but not significantly better) performance at second testing session, irrespective of modality. No significant difference in digit forward scores (differences in test scores across testing modalities not provided)	Administered in Spanish
	Wadsworth (2018)	N = 197, n = 78 MCI/AD, n = 119 Healthy Controls; Mage $_{MCI/AD}$ = 72.71 (SD = 8.43), Meducation $_{MCI/AD}$ = 14.56 (SD = 3.10); 46.2% Female; 61.5% Caucasian, 29.5% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	No main effect of administration modality (ANCOVA controlling for age, education, gender, and depression scores p = .276).	
Digit Span Backwards	Cullum (2014) *	N = 202, n = 83 MCI/AD, n = 119 healthy controls; Mage = 68.5 (SD = 9.5), Meducation = 14.1 (SD = 2.7); 63% Female	H.323 PC - Based Videoconferencing System (Polycom™ iPower 680 Series) that was set up in two non-adjacent rooms	High Speed	Same Day	Similar mean performance across testing modalities. ICC was significant, but fair (.545)	
	Wadsworth (2016) *	N = 84, n = 29 MCI/dementia, n = 55 HC; Mage = 64.89 (SD = 9.73), Meducation = 12.58 (SD = 2.35); 63% Female; 0% Caucasian; 100% American Indian	Polycom iPower 680 series videoconferencing system in same facility	High Speed	Same Day	No differences in mean performance (t = .31, p = .760). ICC = .69 (good)	
	Vahia (2015)	N = 22 spanish-speaking Hispanics living in US who were referred	CODEC (coder-decoder) capable of simultaneously	512 kbit/s	2 weeks	Using mixed-effects models, no significant difference between	Administered in Spanish

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
		for cognitive testing by psychiatrist; Mage = 70.75, Meducation = 5.54 ; 22.75% Female; 0% Caucasian; 100% Hispanic	streaming video and content (i.e., laptop screen) on side by side monitors, remotely controlled Pan Tilt and Zoom cameras, a tablet PC laptop, videoconference microphone; and dual 26 inch LCD TVs			testing modalities on overall Cognitive Composite score derived from entire test battery (F(1,37) = .31, *p* =.0579), with slightly better (but not significantly better) performance at second testing session, irrespective of modality. No significant difference in digit forward scores (differences in test scores across testing modalities not provided)	
	Wadsworth (2018)	N = 197, n = 78 MCI/AD, n = 119 Healthy Controls; Mage $_{MCI/AD}$ = 72.71 (SD = 8.43), Meducation $_{MCI/AD}$ = 14.56 (SD = 3.10); 46.2% Female; 61.5% Caucasian, 29.5% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	No main effect of administration modality (ANCOVA controlling for age, education, gender, and depression scores *p = .635*)	
Digit Span Total	Cullum (2006) *	N = 33, n = MCI, n = AD; Mage = 73.5 (SD = 6.9), Meducation = 15.1 (SD = 2.7); 33% Female; 97% Caucasian	An H.323 PC-based Videoconferencing System was set up in two nonadjacent rooms using a Polycom iPower 680 Series videoconferencing system (2 units)	High Speed	Same Day	Similar mean scores between two testing modalities. ICC = .78 (excellent)	
	Grosch (2015) *	N = 8 geropsychiatric VA patients; 12.5% Female	An H.323 PC-based Videoconferencing System at VAMC	384 kb/s	Same Day	No difference in mean test score across conditions. ICC = .72 (good)	
Brief Test of Attention	Hildebrand (2004) *	N = 29 HC from Canada, Mage = 68 (SD = 8), Meducation = 13 (SD = 3), *MMMSE* = 28.9 (SD = 1.3); 73% Female	Videoconferencing systems (LC5000, VTEL) with two 81cm monitors and far-end camera control. Conducted on-site	336 kbit/s	2–4 weeks apart	No difference in mean score across conditions (Limits of agreeement −5.09 – 6.95)	

Notes.

* indicates a study from Brearly et al., (2017)

Clin Neuropsychol. Author manuscript; available in PMC 2022 March 07.

Table 5.

Tests of Processing Speed

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
Oral Trails A	Wadsworth (2016) *	N = 84, n = 29 MCI/dementia, n = 55 HC; Mage = 64.89 (SD = 9.73), Meducation = 12.58 (SD = 2.35); 63% Female; 0% Caucasian; 100% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	Small but significant difference with better face-to-face performance (t = −9.60, p = < *.001).* ICC = .83 (excellent)	

Notes.

* indicates a study from Brearly et al., (2017)

Clin Neuropsychol. Author manuscript; available in PMC 2022 March 07.

Table 6.

Tests of Language

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
BNT	Vestal (2006) *	N = 10 Memory Disorder Referral at VA; Mage = 73.9 (SD = 3.7), MMMSE = 26.1 (SD = 1.4)	Television monitor and a microphone completed at VA; technician in room to help	384 kbit/s	Same Day	Wilcox signed Rank Test non-significant (z = −0.171, *p* = 0.864)	
BNT-15	Cullum (2014) *	N = 202, n = 83 MCI/AD, n = 119 healthy controls; Mage = 68.5 (SD = 9.5), Meducation = 14.1 (SD = 2.7); 63% Female	H.323 PC - Based Videoconferencing System (Polycom™ iPower 680 Series) that was set up in two non-adjacent rooms	High Speed	Same Day	Similar mean scores between testing modalities. ICC = .812 (excellent)	
	Cullum (2006) *	N = 33, n = MCI, n = AD; Mage = 73.5 (SD = 6.9), Meducation = 15.1 (SD = 2.7); 33% Female; 97% Caucasian	An H.323 PC-based Videoconferencing System was set up in two nonadjacent rooms using a Polycom iPower 680 Series videoconferencing system (2 units)	High Speed	Same Day	Similar mean scores between testing modalities. ICC = .87 (excellent)	
	Wadsworth (2016) *	N = 84, n = 29 MCI/dementia, n = 55 HC; Mage = 64.89 (SD = 9.73), Meducation = 12.58 (SD = 2.35); 63% Female; 0% Caucasian; 100% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	Small but significant difference with better face-to-face performance (*t* = 3.21, *p = .002)*. ICC = .93 (excellent)	
	Wadsworth (2018)	N = 197, n = 78 MCI/AD, n = 119 Healthy Controls; $Mage_{MCI/AD}$ = 72.71 (SD = 8.43), $Meducation_{MCI/AD}$ = 14.56 (SD = 3.10); 46.2% Female; 61.5% Caucasian, 29.5% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	No main effect of administration modality (ANCOVA controlling for age, education, gender, and depression scores). Health controls performed better than MCI/AD	
Ponton-Satz Spanish Naming Test	Vahia (2015)	N = 22 spanish-speaking Hispanics living in US who were referred for cognitive testing by psychiatrist; Mage = 70.75, Meducation = 5.54 ; 22.75% Female; 0% Caucasian; 100% Hispanic	CODEC (coder-decoder) capable of simultaneously streaming video and content (i.e., laptop screen) on side by side monitors, remotely controlled Pan Tilt and Zoom cameras, a tablet PC laptop, videoconference microphone; and dual 26 inch LCD TVs	512 kbit/s	2 weeks	Using mixed-effects models, no significant difference between testing modalities on overall Cognitive Composite score derived from entire test battery (F(1,37) = .31, *p* =.0579), with slightly better (but not significantly better) performance at second testing session, irrespective of modality. No significant difference in naming scores (differences in test scores across testing modalities not provided)	Administered in Spanish

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
Letter Fluency	Cullum (2014) *	N = 202, n = 83 MCI/AD, n = 119 healthy controls; Mage = 68.5 (SD = 9.5), Meducation = 14.1 (SD = 2.7); 63% Female	H.323 PC - Based Videoconferencing System (Polycom™ iPower 680 Series) that was set up in two non-adjacent rooms	High Speed	Same Day	Similar mean scores between testing modalities. ICC = .848 (excellent)	
	Cullum (2006) *	N = 33, n = MCI, n = AD; Mage = 73.5 (SD = 6.9), Meducation = 15.1 (SD = 2.7); 33% Female; 97% Caucasian	An H.323 PC-based Videoconferencing System was set up in two nonadjacent rooms using a Polycom iPower 680 Series videoconferencing system (2 units)	High Speed	Same Day	Similar mean scores between testing modalities. ICC = .83 (excellent).	
	Hildebrand (2004) *	N = 29 HC from Canada, Mage = 68 (SD = 8), Meducation = 13 (SD = 3), *MMMSE* = 28.9 (SD = 1.3); 73% Female	Videoconferencing systems (LC5000, VTEL) with two 81cm monitors and far-end camera control. Conducted on-site	336 kbit/s	2–4 weeks apart	No difference in mean score across conditions (Limits of agreeement −4.34 – 4.82)	
	Vestal (2006) *	N = 10 Memory Disorder Referral at VA; Mage = 73.9 (SD = 3.7), MMMSE = 26.1 (SD = 1.4)	Television monitor and a microphone completed at VA; technician in room to help	384 kbit/s	Same Day	Wilcox signed Rank Test non-significant ($z = -1.316$, $p = 0.188$)	
	Wadsworth (2016) *	N = 84, n = 29 MCI/dementia, n = 55 HC; Mage = 64.89 (SD = 9.73), Meducation = 12.58 (SD = 2.35); 63% Female; 0% Caucasian; 100% American Indian	Polycom iPower 680 series videoconferencing system in same facility	High Speed	Same Day	No differences in mean performance ($t = -0.10$, $p = .920$; Bonferonni correction set alpha to .004). ICC = .93 (excellent)	
	Vahia (2015)	N = 22 spanish-speaking Hispanics living in US who were referred for cognitive testing by psychiatrist; Mage = 70.75, Meducation = 5.54 ; 22.75% Female; 0% Caucasian; 100% Hispanic	CODEC (coder-decoder) capable of simultaneously streaming video and content (i.e., laptop screen) on side by side monitors, remotely controlled Pan Tilt and Zoom cameras, a tablet PC laptop, videoconference microphone; and dual 26 inch LCD TVs	512 kbit/s	2 weeks	Using mixed-effects models, no significant difference between testing modalities on overall Cognitive Composite score derived from entire test battery ($F(1,37) = .31$, $p = .0579$), with slightly better (but not significantly better) performance at second testing session, irrespective of modality. No significant difference in letter fluency performance (differences in test scores across testing modalities not provided)	Administered in Spanish
	Wadsworth (2018)	N = 197, n = 78 MCI/AD, n = 119 Healthy Controls; $Mage_{MCI/AD}$ = 72.71 (SD = 8.43), $Meducation_{MCI/AD}$ = 14.56 (SD = 3.10); 46.2% Female; 61.5% Caucasian, 29.5% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	No main effect of administration modality (ANCOVA controlling for age, education, gender, and depression scores $p = .814$).	

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
Category Fluency	Cullum (2014) *	N = 202, n = 83 MCI/AD, n = 119 healthy controls; Mage = 68.5 (SD = 9.5), Meducation = 14.1 (SD = 2.7); 63% Female	H.323 PC - Based Videoconferencing System (Polycom™ iPower 680 Series) that was set up in two non-adjacent rooms	High Speed	Same Day	Similar mean performance across testing modalities. ICC = .719 (good)	
	Cullum (2006) *	N = 33, n = MCI, n = AD; Mage = 73.5 (SD = 6.9), Meducation = 15.1 (SD = 2.7); 33% Female; 97% Caucasian	An H.323 PC-based Videoconferencing System was set up in two nonadjacent rooms using a Polycom iPower 680 Series videoconferencing system (2 units)	High Speed	Same Day	Similar mean scores between testing modalities. ICC = .58 (fair), which was below threshold of .60.	
	Vahia (2015)	N = 22 spanish-speaking Hispanics living in US who were referred for cognitive testing by psychiatrist; Mage = 70.75, Meducation = 5.54 ; 22.75% Female; 0% Caucasian; 100% Hispanic	CODEC (coder-decoder) capable of simultaneously streaming video and content (i.e., laptop screen) on side by side monitors, remotely controlled Pan Tilt and Zoom cameras, a tablet PC laptop, videoconference microphone; and dual 26 inch LCD TVs	512 kbit/s	2 weeks	Using mixed-effects models, no significant difference between testing modalities on overall Cognitive Composite score derived from entire test battery ($F(1,37) = .31$, $p = .0579$), with slightly better (but not significantly better) performance at second testing session, irrespective of modality. No significant difference in category fluency performance (differences in test scores across testing modalities not provided)	Administered in Spanish
	Wadsworth (2016) *	N = 84, n = 29 MCI/dementia, n = 55 HC; Mage = 64.89 (SD = 9.73), Meducation = 12.58 (SD = 2.35); 63% Female; 0% Caucasian; 100% American Indian	Polycom iPower 680 series videoconferencing system in same facility	High Speed	Same Day	No differences in mean performance ($t = 2.27$, $p = .026$; Bonferonni correction set alpha to .004). ICC = .74 (good)	
	Wadsworth (2018)	N = 197, n = 78 MCI/AD, n = 119 Healthy Controls; $Mage_{MCI/AD}$ = 72.71 (SD = 8.43), $Meducation_{MCI/AD}$ = 14.56 (SD = 3.10); 46.2% Female; 61.5% Caucasian, 29.5% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	Small but significant effect of administration modality ($p < .001$; d = .184 for MCI/AD)	
Token Test	Vestal (2006) *	N = 10 Memory Disorder Referral at VA; Mage = 73.9 (SD = 3.7), MMMSE = 26.1 (SD = 1.4)	Television monitor and a microphone completed at VA; technician in room to help	384 kbit/s	Same Day	Wilcox signed Rank Test non-significant ($z = -1.084$, $p = 0.279$)	. In addition, a clinician-generated Token Test palate (Appendix B) was used to aid the participants in the correct re-formation of the tokens while being

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
							administered the Token Test
Picture Description	Vestal (2006) *	N = 10 Memory Disorder Referral at VA; Mage = 73.9 (SD = 3.7), MMMSE = 26.1 (SD = 1.4)	Television monitor and a microphone completed at VA; technician in room to help	384 kbit/s	Same Day	Wilcox signed Rank Test non-significant (z = 0.0, p = 1.00)	Cookie theft and picnic scenes
Aural Comprehnsion of Words and Phrases	Vestal (2006) *	N = 10 Memory Disorder Referral at VA; Mage = 73.9 (SD = 3.7), MMMSE = 26.1 (SD = 1.4)	Television monitor and a microphone completed at VA; technician in room to help	384 kbit/s	Same Day	Wilcox signed Rank Test non-significant (z = −1.20, p = 0.230)	

Notes.

* indicates a study from Brearly et al., (2017)

Table 7.

Tests of Memory

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
HVLT Immediate	Cullum (2014) *	N = 202, n = 83 MCI/AD, n = 119 healthy controls; Mage = 68.5 (SD = 9.5), Meducation = 14.1 (SD = 2.7); 63% Female	H.323 PC - Based Videoconferencing System (Polycom™ iPower 680 Series) that was set up in two non-adjacent rooms	High Speed	Same Day	Significantly higher test performance for video administration (23.4 (SD = 6.90) vs. 22.5 (SD = 6.98), p = .005). ICC was still excellent (.798)	
	Cullum (2006) *	N = 33, n = MCI, n = AD; Mage = 73.5 (SD = 6.9), Meducation = 15.1 (SD = 2.7); 33% Female; 97% Caucasian	An H.323 PC-based Videoconferencing System was set up in two nonadjacent rooms using a Polycom iPower 680 Series videoconferencing system (2 units)	High Speed	Same Day	Similar mean scores between two testing modalities. ICC = .77 (excellent)	
	Wadsworth (2016) *	N = 84, n = 29 MCI/dementia, n = 55 HC; Mage = 64.89 (SD = 9.73), Meducation = 12.58 (SD = 2.35); 63% Female; 0% Caucasian; 100% American Indian	Polycom iPower 680 series videoconferencing system in same facility	High Speed	Same Day	No differences in mean performance (t = −2.11, p = .038; Bonferonni correction set alpha to .004). ICC = .88 (excellent)	
	Vahia (2015)	N = 22 spanish-speaking Hispanics living in US who were referred for cognitive testing by psychiatrist; Mage = 70.75, Meducation = 5.54 ; 22.75% Female; 0% Caucasian; 100% Hispanic	CODEC (coder-decoder) capable of simultaneously streaming video and content (i.e., laptop screen) on side by side monitors, remotely controlled Pan Tilt and Zoom cameras, a tablet PC laptop, videoconference microphone; and dual 26 inch LCD TVs	512 kbit/s	2 weeks	Using mixed-effects models, no significant difference between testing modalities on overall Cognitive Composite score derived from entire test battery (F(1,37) = .31, p =.0579), with slightly better (but not significantly better) performance at second testing session, irrespective of modality. No significant difference in HVLT scores (differences in test scores across testing modalities not provided).	Administered in Spanish
	Wadsworth (2018)	N = 197, n = 78 MCI/AD, n = 119 Healthy Controls; Mage $_{MCI/AD}$ = 72.71 (SD = 8.43), Meducation $_{MCI/AD}$ = 14.56 (SD = 3.10); 46.2% Female; 61.5% Caucasian, 29.5% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	No main effect of administration modality (ANCOVA controlling for age, education, gender, and depression scores p = .457)	
HVLT Delay	Cullum (2006)	N = 33, n = MCI, n = AD; Mage = 73.5 (SD = 6.9), Meducation = 15.1 (SD = 2.7); 33% Female; 97% Caucasian	An H.323 PC-based Videoconferencing System was set up in two nonadjacent rooms using a Polycom iPower 680 Series videoconferencing system (2 units)	High Speed	Same Day	Slightly higher, but non-significant performance for face-to-face testing. ICC = .61 (good)	

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
	Wadsworth (2016) *	N = 84, n = 29 MCI/dementia, n = 55 HC; Mage = 64.89 (SD = 9.73), Meducation = 12.58 (SD = 2.35); 63% Female; 0% Caucasian; 100% American Indian	Polycom iPower 680 series videoconferencing system in same facility	High Speed	Same Day	No differences in mean performance (t = −2.25, p = .027; Bonferonni correction set alpha to .004). ICC = .90 (excellent)	
	Wadsworth (2018)	N = 197, n = 78 MCI/AD, n = 119 Healthy Controls; Mage $_{MCI/AD}$ = 72.71 (SD = 8.43), Meducation $_{MCI/AD}$ = 14.56 (SD = 3.10); 46.2% Female; 61.5% Caucasian, 29.5% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	No main effect of administration modality (ANCOVA controlling for age, education, gender, and depression scores p = .735)	
BVMT-R	Vahia (2015)	N = 22 spanish-speaking Hispanics living in US who were referred for cognitive testing by psychiatrist; Mage = 70.75, Meducation = 5.54 ; 22.75% Female; 0% Caucasian; 100% Hispanic	CODEC (coder-decoder) capable of simultaneously streaming video and content (i.e., laptop screen) on side by side monitors, remotely controlled Pan Tilt and Zoom cameras, a tablet PC laptop, videoconference microphone; and dual 26 inch LCD TVs	512 kbit/s	2 weeks	Using mixed-effects models, no significant difference between testing modalities on overall Cognitive Composite score derived from entire test battery (F(1,37) = .31, p =.0579), with slightly better (but not significantly better) performance at second testing session, irrespective of modality. No significant difference in BVMT performance (differences in test scores across testing modalities not provided).	Administered in Spanish

Notes.

* indicates a study from Brearly et al., (2017)

Clin Neuropsychol. Author manuscript; available in PMC 2022 March 07.

Table 8.

Tests of Executive Functioning

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
Clock Drawing Test	Cullum (2014) *	N = 202, n = 83 MCI/AD, n = 119 healthy controls; Mage = 68.5 (SD = 9.5), Meducation = 14.1 (SD = 2.7); 63% Female	H.323 PC - Based Videoconferencing System (Polycom™ iPower 680 Series) that was set up in two non-adjacent rooms	High Speed	Same Day	Similar mean performance. ICC = .709 (moderate)	The visuospatial measures were scored by asking participants to hold up their paper in front of camera
	Cullum (2006) *	N = 33, n = MCI, n = AD; Mage = 73.5 (SD = 6.9), Meducation = 15.1 (SD = 2.7); 33% Female; 97% Caucasian	An H.323 PC-based Videoconferencing System was set up in two nonadjacent rooms using a Polycom iPower 680 Series videoconferencing system (2 units)	High Speed	Same Day	Kappa was moderate (0.48, $p < .0001$)	The visuospatial measures were scored via the television monitor by asking participants to hold up their paper in front of camera
	Wadsworth (2016) *	N = 84, n = 29 MCI/dementia, n = 55 HC; Mage = 64.89 (SD = 9.73), Meducation = 12.58 (SD = 2.35); 63% Female; 0% Caucasian; 100% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	No differences in mean performance. ICC = .65 (good)	objects were asked to holdup their drawing to the camera in order for the examiner to score it immediately
	Wadsworth (2018)	N = 197, n = 78 MCI/AD, n = 119 Healthy Controls; $Mage_{MCI/AD}$ = 72.71 (SD = 8.43), $Meducation_{MCI/AD}$ = 14.56 (SD = 3.10); 46.2% Female; 61.5% Caucasian, 29.5% American Indian	Polycom iPower 680 series videoconferencing system	High Speed	Same Day	No main effect of administration modality (ANCOVA controlling for age, education, gender, and depression scores $p = .520$)	
	Grosch (2015) *	N = 8 geropsychiatric VA patients; 12.5% Female	An H.323 PC-based Videoconferencing System at VAMC	384 kb/s	Same Day	No difference in mean test score across conditions. ICC was non-significant and just-above “poor” (ICC =.42)	Held up clock to camera for scoring; alterante times used for test-retest
	Vahia (2015)	N = 22 spanish-speaking Hispanics living in US who were referred for cognitive testing by psychiatrist; Mage = 70.75, Meducation = 5.54 ; 22.75% Female; 0% Caucasian; 100% Hispanic	CODEC (coder-decoder) capable of simultaneously streaming video and content (i.e., laptop screen) on side by side monitors, remotely controlled Pan Tilt and Zoom cameras, a tablet PC laptop, videoconference microphone; and dual 26 inch LCD TVs	512 kbit/s	2 weeks	Using mixed-effects models, no significant difference between testing modalities on overall Cognitive Composite score derived from entire test battery (F(1,37) =clock drawing performance (differences in test scores across testing modalities not provided)	Administered in Spanish
	Hildebrand (2004) *	N = 29 HC from Canada, Mage = 68 (SD = 8), Meducation = 13 (SD = 3), *MMMSE* = 28.9 (SD = 1.3); 73% Female	Videoconferencing systems (LC5000, VTEL) with two 81cm monitors and far-end camera control. Conducted on-site	336 kbit/s	2–4 weeks apart	Large, but not significant difference in mean score across conditions (Mean difference −1.93; 95% CI = −5.76 – 1.90); Large Limits	

Clin Neuropsychol. Author manuscript; available in PMC 2022 March 07.

Test Name	Study	Population Characteristics	Video Modality	Connection Speed	Delay Between Assessments	Findings	Administration Notes
						of Agreement (−22.07 – 18.21)	
	Montani (1997) *	N = 14 Mixed rehabilitation; Mage = 88 (SD = 5), 46.7% Female	Camera, television screen, and microphone in an adjacent room	Coaxial	8 days	Large, but non-significant diifference in mean scores across conditions (22.4 vs. 19.8) with better FTF performance. Correlation of test scores significant (r = 0.55)	
Oral Trails B	Wadsworth (2016) *	N = 84, n = 29 MCI/dementia, n = 55 HC; Mage = 64.89 (SD = 9.73), Meducation = 12.58 (SD = 2.35); 63% Female; 0% Caucasian; 100% American Indian	Polycom iPower 680 series videoconferencing system in same facility	High Speed	Same Day	No differences in mean performance (t = −.35, p = .726). ICC = .79 (excellent)	

Notes.

* indicates a study from Brearly et al., (2017)

Clin Neuropsychol. Author manuscript; available in PMC 2022 March 07.