View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by National-Louis University: OASIS - The NLU Digital Commons

Remote Neuropsychological Assessment: Does Mode of Administration Matter?

Renee DeMatteis, M.A.

Florida School of Professional Psychology at National Louis University

Elizabeth Lane, Ph.D.
Chair

Christina D. Brown, Psy.D.
Member

A Clinical Research Project submitted to the Faculty of the Florida School of Professional Psychology at Argosy University in partial fulfillment of the requirementsfor the degree of Doctor of Psychology in Clinical Psychology

Tampa, Florida

April, 2021

The Doctorate Program in Clinical Psychology
Florida School of Professional Psychology
at National Louis University

CERTIFICATE OF APPROVAL

Clinical Research Project

This is to certify that the Clinical Research Project of

Renee DeMatteis, M.A.

has been approved by the
CRP Committee on May 12, 2021
as satisfactory for the CRP requirement
for the Doctorate of Psychology degree
with a major in Clinical Psychology

Examining Committee:

Elizabeth M. Lane, PhD
Committee Chair: Elizabeth Lane, Ph.D.

Christina D Brown Psy.D.
Member: Christina Brown, Psy.D.

ABSTRACT

Neuropsychology relies heavily on standardization of administration to increase the validity, reliability, sensitivity, and specificity of an assessment instrument. The COVID-19 pandemic has rapidly increased the need to be able to provide neuropsychological assessments remotely. Teleneuropsychology can be conducted through various avenues, including telephone, computerized, and televideo modalities. Given neuropsychology's reliance on standardization for proper use of normative data that accompanies individual assessments, the question arises how mode of administration impacts the validity, reliability, sensitivity, and specificity of an assessment instrument. The literature review summarizes the research conducted regarding the validity, reliability, sensitivity, and specificity of telephone neuropsychological assessments, computerized neuropsychological assessments, and televideo neuropsychological assessments. Additionally, the literature review aims to provide guidelines for best practice for each mode of administration for practicing neuropsychologists.

REMOTE NEUROPSYCHOLOGICAL ASSESSMENT: DOES MODE OF ADMINISTRATION MATTER?

Copyright ©2021

Renee DeMatteis, M.A.

All rights reserved

DEDICATION

I dedicate this manuscript to my parents, James and Dorie DeMatteis, who provided me with unconditional love and support and taught me to always believe in myself.

ACKNOWLEDGEMENTS

Nothing great is ever accomplished alone. First, I would like to acknowledge both my committee chair, Dr. E. Lane, and committee member, Dr. C. Brown, for their continued guidance and support over the years. Thank you for not only being role models but for being my mentors. You both have inspired me to always take on a challenge and to push forward despite adversity.

To my parents, James and Dorie DeMatteis, thank you for believing in me and supporting me always. Thank you for your encouragement and guidance throughout graduate school and the clinical research project process. You always taught me push forward even in the face of adversity. Lastly, thank you for your unconditional love and support. I would not be here without you both.

To my closest friends, Lacy Davis and Zandra McDonald, I have no idea how I could have made it this far without you both. Thank you for all the late-night conversations and always laughing with me when needed. Your friendship and encouragement are second to none.

Lastly, to the psychologists at Northshore Psychological Association, thank you for your continued supervision and guidance throughout the past few years. Thank you for your encouragement through this process and to always move forward.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I: INTRODUCTION

Neuropsychology, as defined by Lezak et al. (2012), is studying the relationship between the brain and behavior. The purpose of neuropsychology is to integrate information from a multidisciplinary viewpoint. Neuropsychologists use the information provided by neurology, psychiatry, biology, pharmacology, psychology, and physiological psychology to understand better the relationship between the brain and behavior and how that relates to cognitive deficits. Telemedicine is delivering health-related services through electronic communications (Grosch et al., 2011). Teleneuropsychology is a subgroup of telemedicine that uses electronic communications to administer neuropsychological assessments. Modalities of electronic communications may include telephone, video conferencing, virtual reality, email, and wireless phones (Grosch et al., 2011).

Neuropsychology relies heavily on standardization of administration to increase the validity, reliability, sensitivity, and specificity of an assessment. Psychometric properties of neuropsychological assessments are important for understanding and interpreting test results to make informed and accurate clinical judgments and diagnoses (Brooks et al., 2009). A study conducted by Shum et al. (1997) evaluated the speed of the presentation of logical memory in the WAIS-R on a person's ability to recall the information. These researchers found that participants in the fast group recalled less story information, thus indicating the speed at which an examiner reads the information can impact a person's ability to process and recall information (Shum et al., 1997). This is just one example of how deviation from standardization can impact test results.

Standardization

Neuropsychology depends on the standardization of assessment to have reliable normative data. If we do not have reliable normative data, then it is near impossible to determine if an individual is displaying cognitive deficits that are outside of the expectations given their demographic information (age, education, gender, and ethnicity). The primary goal of neuropsychological evaluations is to assess an individual's current level of cognitive functioning; as such, accuracy is extremely important. When evaluations deviate from standardization, this can impact the accuracy of test results because normative data may not be the most reliable given the administration changes. As such, inaccurate test results can lead to improper diagnosis and treatment recommendations thereby providing further evidence as to why standardization is extremely important. Standard and optimal conditions should both be considered during test administration and selection.

Standard Test Conditions

Each test developer intended a specific set of conditions, which is considered test standardization or standard conditions that inform examiners on test administration and scoring procedures. These standard conditions for each neuropsychological assessment were utilized to accurately norm the assessment that provides an accurate and comparable score to determine if there is a deficit or impairment. Therefore, when there is a deviation from standardization, this calls into question the validity of the measure and if it has been compromised (AERA, APA, & NCME, 2014). Neuropsychological assessments come with an administration manual that outlines how the test should be administered. This administration protocol matches how the tests were normed.

Wechsler (2008), Wechsler (2009), and Mather and Woodcock (2001) all have similar variations of how the physical environment should be when administering their neuropsychological assessments. To provide an ideal testing environment, the examiner should administer the test in a well-lit, quiet room, and the room should be free from distractions and interruptions. To help the examinee focus, external distractions during test administration should be minimized. These assessments ask for specific seating arrangements, which, if not followed, reduce the assessment's validity. The examiner should sit directly across from the examinee so the examiner can fully observe his or her behavior and responses, and the examinee should be seated at a table or desk. This study wants to address the impact of technology on the administration of neuropsychological assessments. Many neuropsychological assessments provide specific stimulus books, materials, and time measurements that can be impacted by videoconferencing. Videoconferencing and computerized test administration utilize the Internet although there have been significant advances in this technology. It is still fallible and can impact timed measures and the proper delivery of test instructions (Bilder et al., 2020).

Optimal Test Conditions

Optimal test conditions are used frequently in psychological practice as these conditions help an examinee to perform their best. Optimal test conditions take into consideration factors that can impact cognitive performance, such as fatigue, distraction, and test anxiety. As such, it is important to provide a private and comfortable setting that limits distractions. In order to address test anxiety, it is important to build rapport with each examinee in an attempt to offer a benevolent emotional environment. For example, an examiner can adhere to test instructions as well as not giving hints regarding the accuracy of response while providing a climate that does not create fear or discomfort as the examiner can gently encourage the examinee in a way that makes them feel more comfortable thus allowing them to perform at their best. An ideal way to reduce test fatigue is to offer breaks during test administration and or take into consideration the examinee limits to split testing into multiple days because fatigue can be a chronic problem in many neurological conditions (Lezak et al., 2012).

Addressing Deviation from Standardization

According to APA standards for psychological testing, when testing conditions deviate from standardization established by test developers, this should be identified, explained, and documented to both the patient and in the neuropsychological report. Additional standards set out by the interorganizational practice committee for teleneuropsychology indicate that the provider must gain informed consent for teleneuropsychology practice (Bilder et al., 2020). There may be a reduction of the validity of the scores when measures are administered under alternative conditions or deviate from standardization (AERA et al., 2014). As described above, neuropsychological evaluations are often administered under varying test conditions. In order to maintain an ethical practice, the provider must describe the deviations that occurred from standardization, the limitations of the test results, and the impact on diagnostic conclusions and treatment recommendations (AERA et al., 2014; Bilder et al., 2020).

Current Guidelines for Telepsychology

The American Psychological Association [APA] (2013) created a set of guidelines to help psychologists practice telepsychology. Specifically, a guideline was created to help address assessment administration via telehealth. "Guideline 7. Psychologists are encouraged to consider the unique issues that may arise with test instruments and assessment approaches designed for in-person implementation when providing telepsychology services" (APA, 2013, p. 798). APA (2013) states the purpose of this guideline is to help address the deviation from standardization in that most neuropsychological assessments were specifically designed for an in-person assessment. Furthermore, they encourage psychologists to be aware of the impacts this deviation can have on properly administering and interpreting these assessments when procedures are changed to be conducted via telehealth. In regard to test administration, this guideline specifically addresses the need to have suitable psychometric properties (e.g., reliability and validity) to ensure test integrity is preserved when assessments are adapted for telehealth administration (APA, 2013). APA (2013) also addresses the need to ensure quality technology and the equipment requirements to properly conduct assessment via telehealth. Additionally, they discuss the need for the psychologist to be aware and ready to address the differences between results obtained in person and via telehealth. Psychologists are also encouraged to properly document test procedure adaptations or modifications made as well as the results from the assessment. Lastly, APA (2013) addresses the need for proper test norms when using telehealth when available. Essentially, it is of the utmost importance that the psychologist strives to use norms that were created from telehealth administration when available; however, if those are not available, the psychologist should address the limitations of the assessment procedure and norms.

The Interorganizational Practice Committee for Teleneuropsychology provides additional concerns and recommendations regarding the practice of teleneuropsychology. Specifically, they encourage providers to address limitations in current research regarding the practice of teleneuropsychology as many tests have been considered valid in the administration of videoconferencing; however, the impact of test results in the reduction of confidence in the diagnostic conclusions and the impact of treatment recommendations is not well-known (Bilder et al., 2020). Furthermore, the provider should address the loss of qualitative data, which is usually obtained during in-person exams, and how this will further limit conclusions and recommendations. Bilder et al. (2020) discuss the need to address these concerns in both the informed consent and written test results.

Hewitt and Loring (2020) wrote a review on their clinic at Emory addressing how they transferred to a telehealth clinic during COVID-19. Hewitt and Loring (2020) addressed many aspects of a teleneuropsychological clinic that should be considered when practicing teleneuropsychology. Some of these aspects included updated informed consent, addressing appropriate patients, test modifications, and documentation. Hewitt and Loring (2020) discussed the need for appropriate informed consent, which goes beyond in-person informed consent. Specifically, APA (2013) discusses in guideline 3,

> "Psychologists strive to obtain and document informed consent that specifically addresses the unique concerns related to the telepsychology services they provide. When doing so, psychologists are cognizant of the applicable laws and regulations, as well as organizational requirements, that govern informed consent in this area." (p. 795)

In regard to appropriate patients, Hewitt and Loring (2020) did not see any legal cases or epilepsy surgery cases. Additionally, the clinic assessed patients' ability to use technology and patients' own comfort level to address whether a patient is appropriate for teleneuropsychology.

Factors Impacting Cognitive Functioning and Test Assessment

According to Lezak et al. (2012), there are common assessment problems with brain disorders and administering neuropsychological assessments in hospitals, some of which are fatigue, medication, and pain. Often in the hospital setting, a patient may be experiencing any of these. Patients with brain disorders tend to fatigue easily, especially when an acute condition happened recently, such as experiencing a stroke, traumatic brain injury, cancer, chemotherapy, and respiratory disease. Fatigue can complicate neuropsychological testing because it impacts many cognitive domains including sustained attention, concentration, reaction time, and processing speed. Studies of sleep deprivation have found complications in the cognitive domains of psychomotor vigilance, executive function, and psychomotor speed and accuracy (Lezak et al., 2012).

Medication is often changed while a patient is in the hospital that can cause complications with a person's cognitive functioning. The neuropsychologist needs to understand the origin of the deficit (i.e., is it caused by something organic or environmental in nature). Medication is shown to have varying effects on cognitive function due to the many medications that patients receive; however, most deficits are seen with anticholinergics, benzodiazepines, narcotics, neuroleptics, antiepileptic drugs, and sedative-hypnotics. New medications or changes in medications often can cause changes in mental efficacy for a few weeks. Chemotherapy has been linked to cognitive dysfunction, including difficulties with concentration and memory (Lezak et al., 2012).

Pain syndromes are common among the general population. However, chronic pain is often a comorbid symptom with TBI, brain disorders including thalamic stroke, multiple sclerosis, or disease involving cranial or peripheral nerves. Pain can impact attentional capacity, processing speed, and psychomotor speed. Studies looking at TBI with pain and without pain found that those with pain tending to perform more poorly included difficulties with learning and problem-solving.

Many studies have looked at how bed rest or physical inactivity can impact cognitive functioning. Lipnicki and Gunga (2008) reviewed results from bed-rest studies and found that bed rest only, excluding head-down tilt bed rest, has a slower reaction time after bed rest, ranging from seven to 70 days. Other implications of bed rest include worsening mental arithmetic abilities, short-term memory, and executive function. Another study examined motivating factors for exercise during a hospital stay. Many patients cited the negative effects of prolonged bed rest as the primary motivator for exercise (So & Pierluissi, 2012). The negative effects included pain, fatigue, and short-term memory difficulties. Lastly, a study done by Lipnicki et al. (2009) examined executive functioning changes in healthy males after 60 days of bed rest. They found changes in the prefrontal cortex that relate to executive functioning deficits and a slower reaction time.

Additionally, when providing remote neuropsychological assessments, some impacts should be addressed and understood that differ from in-person evaluations and can impact the interpretation of test results on diagnostic conclusions. Bilder et al. (2020)

address bandwidth concerns and considerations that can impact the administration of assessments. As such, providers are encouraged to track and document technical problems such as disconnection or lag in a video (Bilder et al., 2020). Additionally, it is important to understand patients' individual comfort level with technology and their familiarity with online platforms. If a patient has limited familiarity with online platforms, this may increase test anxiety and should be addressed in the report (Bilder et al., 2020). Additional distractions in the home where the examinee is located, such as family members or caregivers, may impact the patient's level of distraction and/or anxiety. Again, the provider should track and document any interruptions and distractions including sounds, family members, and/or pets walking in (Bilder et al., 2020). The provider should also consider the impact on their ability to build rapport in remote settings compared with typical social communication in that teleneuropsychology may impact a provider's ability to discern data from body language, facial expression, and tone of voice (Bilder et al., 2020). Lastly, neuropsychological evaluations rely heavily on behavioral observations to make diagnostic conclusions. According to Bilder et al. (2020), behavioral observations can be impacted when assessments are administered remotely, as such, the provider should be aware that there may be a loss of some qualitative data that can affect the clinical understanding and limit conclusions and recommendations.

Considering the impacts of fatigue, medication, pain, and bed rest on a patient's cognitive functioning, an examiner needs to understand the complications that can arise when testing in a hospital bed or remotely. The provider also needs to be aware of how remote assessments can impact the quality of data received and impact the examinees' fatigue, anxiety, and distractibility on clinical conclusions. Because these factors can all be applied to a patient in a hospital or remote settings, one must know if impairments in cognitive functioning are due to organic deficits, complications from an atypical testing environment, or these factors.

Current Clinical Research Project

In many clinical settings, deviation from standardization occurs for many reasons. Currently, the use of teleneuropsychology has increased due to the global pandemic COVID-19. Teleneuropsychology has been used for many reasons including rural settings, people with insufficient healthcare resources in their community, individuals with disabilities which limited mobility, and victims of natural disasters (Grosch et al., 2011). COVID-19 has disrupted the usual face-to-face contact that is typically utilized in the conventional neuropsychological evaluation. Given COVID-19, in an attempt to maintain a social distancing standard, there has been an increase in the use of teleneuropsychology in order to uphold safety measures for both patients and providers in that older adults are at high risk for contracting COVID-19. Furthermore, the provider may use telephone assessments, videoconferencing assessments, and or self-administered computerized tests to adhere to current social distancing guidelines, which may impact test results (Bilder et al., 2020). Much of the current research addressing teleneuropsychology is done in a controlled environment where patients are seen in a telehealth clinic. Additionally, deviation from standardization also occurs for many reasons during an in-person assessment, primarily in hospital settings due to patients' physical limitations or inability to leave the hospital bed. While providers may request a private room for testing, this is often not the case, as the patient may be incapable of leaving the hospital bed or a testing room may not be available. Therefore, the provider must make do without having a table for the test administration and may use either a hospital bedside tray table or clipboard that may impact timed performance tests and tests involving motor dexterity. The use of clipboards, hospital bedside tray table, telephone, videoconferencing, and self-administered computerized test can violate test standardization, which can invalidate the results. Atypical administration procedures may make it difficult to use the norms derived from standardization and found within assessment manuals.

There are times when a neuropsychological assessment is needed, even when deviance from traditional testing conditions is required. In cases when a deviation is required, it is ethical to continue the evaluation as long as the provider describes the limitation of test results and how diagnostic conclusions derived from the interpretation of test results may be impacted (APA, 2010). As previously mentioned, it is also imperative that the practitioner describe how the testing environment differed from what the test developer intended and indicate that the results should be interpreted with caution because of the test administration differences. Furthermore, in cases with telehealth assessments, there are many pieces of important information that should be addressed with the patient and reports. The goal of this research project is to help address those pieces of information for practicing neuropsychologists.

Purpose of the Literature Review

The purpose of this paper is to summarize the research being conducted on neuropsychological assessments when they are deviating from standardization. This paper will explore current research on teleneuropsychology assessments under varying telecommunication conditions, specifically, telephone assessments, videoconferencing assessments, and self-administered computerized assessments. Additionally, the purpose of this literature review is to better understand how the deviation from standardization impacts diagnostic conclusions and normative data provided in test developers' manuals. Knowledge of test administration and cognitive deficits for neurological diagnoses are used to integrate the research to allow neuropsychologists to use this information to better inform test selection for teleneuropsychology and clinical considerations to take into account test interpretation leading to conclusions and treatment recommendations. Furthermore, this literature review aims to help neuropsychologists better understand the aspects to best practice teleneuropsychology among varying different avenues and what populations are best suited for the practice of teleneuropsychology.

What is telephone cognitive assessments? What assessments have been researched, developed, and/or modified to be administered over the phone? What is the impact on clinical data gathered over the phone, and how it impacts diagnostic conclusions? What populations is it best suited for? It is important to understand the validity of this research and how this is addressed not only in research settings but in clinical settings. What does the neuropsychologist need to know about telephone cognitive assessments for best practice? Lastly, what new research is currently underway in regard to smartphone cognitive applications?

What are computerized neuropsychological assessments? What traditional assessments have been researched, developed, and/or modified to be administered via the computer or a web-based platform? How does this impact current normative data when compared with in-person assessments? What considerations are addressed when providing these assessments and the impact they have on clinical conclusions? What does the neuropsychologist need to provide these assessments in their clinical practice? Lastly, what new research is currently underway regarding the use of computer technology to assess cognitive functioning?

What is televideo cognitive assessments? What assessments have been researched, developed, and/or modified to be administered through video conferencing and their validity in relation to in-person assessments? What impact does televideo assessments have on current normative data? What considerations need to be addressed when providing televideo cognitive assessments in different populations? How is televideo used in clinical practice, and what are the limitations? What information needs to be addressed with patients and documented? Lastly, what new research is currently underway in regard to the use of televideo assessments and their ability to assess cognitive functioning?

Inclusion and Exclusion Criteria

In order to effectively identify the research pertinent to the topic of this literature review, the researcher adhered to specific inclusion and exclusion criteria.

Inclusion

The search engines included Google Scholar, PsycARTICLES, PsycINFO, and Psychology and Behavioral Sciences Collection (EBSCO) using the following search terms: telemedicine, teleneuropsychology, computerized neuropsychological assessment, telecognitive assessment, telephone screening, smartphone cognitive assessments, telephone cognitive assessment, and remote neuropsychological assessments. Research on the validity and reliability of teleneuropsychological assessments with different populations were included. This paper included peer-reviewed literature from the last 30 years (1990-2020). Both original research studies and meta-analyses were included.

Exclusion

For the purpose of this paper, literature reviews and literature not written in English were excluded. Research that did not focus on the validity and reliability of teleneuropsychology were excluded. The research was excluded that focuses on the validity of psychological assessments and teletherapy. Additionally, books written on the teleneuropsychology were not used, and general research reviews were not used.

CHAPTER II: TELEHEALTH: TELEPHONE NEUROPSYCHOLOGICAL ASSESSMENTS

Neuropsychological assessments can be administered in various ways. There have been many research studies conducted into the various different ways neuropsychological assessments can be administered. Interestingly, there has been research into administration of neuropsychological assessments and neuropsychological screeners over the telephone since the late 1980s. More recently, there has been developments in telephone neuropsychological assessment administration given the COVID-19 pandemic. This chapter outlines the various research on cognitive screeners and neuropsychological assessment batteries. Additionally, this chapter provides neuropsychologists pertinent information on populations best suited for telephone assessments, technology needed, and considerations for diagnosis and report writing.

Overview of Current Research

Cognitive Screeners

The telephone interview for cognitive status (TICS) was developed by Brandt et al. (1988). The TICS has a maximum score of 41 and it includes 11 items. These 11 items assess for a person's ability to state their full name, date, address, their ability to count backwards, learn a word list, their ability to subtract, their ability for responsive naming, their ability for word repetition, their ability to name the president/vice president (last name only), finger tapping, and word opposites. In order to assess for finger tapping, the patient was asked to tap their finger five times one the part of the phone they speak into. Brandt et al. (1988) compared 100 Alzheimer's Disease (AD) patients via the telephone with 33 normal control participants in order to examine test-retest reliability.

Additionally, these researchers also compared scores between both the AD group and the normal control group and found a significant difference (t = 15.07, df = 131, p < 0.001) with the AD group scores being lower than the control group. Additionally, the TICS had a strong correlation with the with the Mini Mental Status Examination (MMSE) r = +.95, p < 0.0001; Brandt et al., 1998). Brandt et al. (1988) were able to determine cut off scores via a comparison of mean scores with the TICS and MMSE. The TICS has a cut off score of 31, meaning if a patient has a score of 31 or higher, they are considered “normal” and a score of 30 or lower, they are considered “cognitively impaired.” Based on the analysis, Brandt et al. (1988) found the TICS has a 94% sensitivity and a 100% specificity. Lastly, Brandt et al.’s (1988) research found test-retest reliability (r = +.965, t = 20.82, df = 32 p < 0.0001).

Since the development of the TICS, additional research has been conducted in order to expand on previous research and provide information on its useability with other populations. Rankin et al. (2005) desired to determine if the Age-Related Eye Disease Study (AREDS) could substitute a telephone battery with their in-clinic neuropsychological battery. This study included 1,738 participants with a mean age of 75 years (61 to 87 years) and 57% were female (Rankin et al., 2005). Rankin et al. (2005) compared the Telephone Interview for Cognitive Status – Modified version (TICS-M) with the MMSE. The TICS-M has 50 points and assesses the patient for the ability to state their name, provide the date, provide their age and phone number, counting backwards, word list learning, subtraction, responsive naming, repetition, president/vice president (first and last name), finger tapping, word opposites, and delayed word recall. Rankin et al. (2005) compared scores from the MMSE with the TICS-M and found a weak positive correlation when scores were unadjusted (*p* = 0.44, 95% CI; 0.40 - 0.49). When the scores were adjusted for age and depression at the time of administration, there was a significant positive correlation (*p* = 0.89, 95% CI; 0.88 - 0.90). This indicates that scores between the MMSE and TICS-M are comparable when holding age and depression symptoms constant.

A similar study was conducted by Rapp et al. (2012) to determine the validity of the administration of a neuropsychological battery by phone. This study included 110 female participants ages 65 to 90 years all of whom received both the telephone neuropsychological test battery and in person neuropsychological test battery administered six months apart (Rapp et al., 2012). Rapp et al. (2012) modified the TICS to take out the word list learning task as to avoid proactive interference. Rapp et al. (2012) found no significant difference between telephone (28.8 (2.60)) and in person (29.0 (1.9)) assessments (*p* = 0.71).

Lastly, a study was conducted by Fong et al. (2009) to compare the MMSE with both the TICS-30 and TICS-40 to derive cut off scores. Fong et al. (2009) conducted a longitudinal study that included 746 community dwelling older adults who were gathered from the Aging, Demographics, and Memory Study. These adults were aged 70 to 102 years old with ranging cognition from normal to dementia due to AD, vascular disease, subcortical dementia, frontal lobe dementia, and diffuse Lewy body disease. The TICS-30 has 30 points and assesses for the patient's ability to recall the date, their address, their ability to count backwards, their ability to learn a list of words, ability to complete serial substructions, responsive naming, and word repetition. The TICS-40 is similar to the TICS-30 but adds an additional 10 for delayed word recall of the word list. This study found a mean of 17 (6) for TICS-30 and a mean of 21 (9) for TICS-40 (Fong et al., 2009). When comparing the MMSE with the TICS-30, there was a high correlation with an intraclass correlation coefficient of 0.80 (95% CI; 0.70 - 0.83; Fong et al., 2009). Additionally, there was a high correlation for the MMSE and TICS-40, intraclass correlation coefficient of 0.80 (95% CI; 0.70 - 0.83; Fong et al., 2009). In order derive a cut point, a correlation was calculated for the MMSE cut point and corresponding cut points for the TICS-30 and TICS-40 with a kappa value of 0.69 for both. As such, scores from 25 - 30 on the TICS-30 and 32 - 40 on the TICS-40 is similar to the score 30 on the MMSE.

Wong et al. (2015) examined the 5-minute protocol for the Montreal Cognitive Assessment (MoCA) for telephone administration. One hundred and four patients with stroke or TIA were included in the study to compare mean differences between the MoCA and the MoCA 5-minute protocol (Wong et al., 2015). Half of the participants had cognitive impairment and the groups were comparable in age, education, gender, and stroke severity (Wong et al., 2015). The MoCA 5-minute protocol consists of four subtests assessing for attention, verbal learning and memory, executive functioning/language, and orientation. Modifications included using the number of words recalled in the first immediate recall of the 5-word list in order to measure immediate auditory attention. The study was conducted in Hong Kong; as such, the researchers did not use semantic fluency because the Cantonese language is non-alphabetic. The MoCA 5-minute protocol scores can range from 0-30 (Wong et al., 2015). For the test administration over the phone, participants were asked to turn off the radio or television and go into a quiet room (Wong et al., 2015). When possible, family members were asked to remove distractions and aids such as calendars (Wong et al., 2015). Wong et al. (2015) compared the MoCA with MoCA 5-minute protocol and found they were highly correlated ($r = 0.87$, $p < 0.001$). Additionally, the MoCA 5-minute protocol was able to differentiate between patients with cognitive impairment from those without (AUC for MoCA 5-min protocol = 0.78; MoCA = 0.74, $p > 0.05$ for difference; Cohen's d for group difference 0.8) when compared to the MoCA (Wong et al., 2015). The MoCA 5-minute protocol was equally able to differentiate between those with cognitive impairment in the executive domain from those without (AUC = 0.89, $p < 0.001$; Cohen's d = 1.7 for group difference).

A similar study conducted by Pendlebury et al. (2013) compared 91 non-demented older adults after a cerebral vascular event who initially completed an in-person neuropsychological battery and MoCA with the telephone MoCA (22 points) and short telephone MoCA (verbal fluency, recall, and orientation; 12 points): only 73 participants completed the telephone version of testing one month after initial face-to-face testing. Modifications made during the telephone MoCA included having participants tap the side of the telephone with a pencil for the sustained attention task instead of tapping the desk during face-to-face administration (Pendlebury et al., 2013). Of note, these researchers did not add an additional point for low education during telephone administration; however, it was added in face-to-face administration. Pendlebury et al. (2013) found worse scores on MoCA repetition, abstraction, and verbal fluency on telephone versus face-to-face administration ($p < 0.02$) even when excluding patients with hearing difficulties. In regard to telephone administration's reliability to diagnose mild cognitive impairment for single domain, T-MoCA was 0.75 (95% CI, 0.64

- 0.87) and T-MoCA short was 0.72 (95% CI, 0.60 - 0.84; Pendlebury et al., 2013). However, for multiple domain MCI, the reliability of telephone administration increased as T-MoCA was 0.85 (95% CI, 0.75 - 0.96) and was 0.85 for the T-MoCA short (95% CI, 0.75 - 0.96; Pendlebury et al., 2013). Pendlebury et al. (2013) derived cutoff scores for optimal sensitivity and specificity for both T-MoCA (18 to 19) and T-MoCA short (10 to 11). In conclusion, these researchers found that the T-MoCA is a valid test for assessing cognition. However, one must be aware that some subtests can be negatively impacted by telephone administration specifically with abstractions, verbal fluency, and repetition. Additionally, these researchers found that the T-MoCA short was worse in detection of single domain MCI. Limitations to note are participants only had a relatively mild cerebral vascular events; as such, consideration for telephone testing may be more difficult with patients with more severe strokes or cognitive impairments. Additionally, this study included a small sample size and an even smaller sample of participants were administered telephone testing. As such, larger studies are needed.

The Mini Mental Status Examination was developed by Folstein et al. (1975). Roccaforte et al. (1992) compared 100 older adult participants from an outpatient geriatric assessment center on both the telephone version of the MMSE and the face-to-face version of the MMSE. Both versions of the MMSE correlated strongly with each other for all participants (Pearson's $r = 0.85$, $p = 0.001$; Roccaforte et al., 1992). Additionally, both these tests were significant for people who had no cognitive impairment ($p = 0.02$) and possible cognitive impairment ($p = 0.002$), mild cognitive impairment ($p = 0.0001$), and moderate cognitive impairment ($p = 0.003$; Roccaforte et al., 1992). Monteiro et al. (1998) evaluated the reliability of the MSE for the use of telephone administration in the assessment of Alzheimer's disease. These researchers compared 30 subjects who were assessed at two different time periods and included 17 females and 13 males (Monteiro et al., 1998). Modifications made to the telephone MMSE are as follows: for naming objects instead of asking the subject to name objects, the examiner asked the participants to name objects they were holding; they also asked questions regarding a watch, such as “What do you wear on your wrist to tell time?” (Monteiro et al., 1998). Additionally, they used caregiver assistance for the three stage commands and writing a sentence as they had the caregiver assist in judging the appropriateness and ability to carry out the command (Monteiro et al., 1998). The researchers did not include the examination of praxis; as such, the total score of the telephone MMSE was out of 29 points. Telephone interclass correlation coefficients for interrater and same in clinic reliability were ICC = 0.98 (Monteiro et al., 1998). Although the correlation coefficients were significant, there are many limitations in this study, specifically the small sample size and limited information regarding new cut offs for modifications made to telephone MMSE.

Telephone Batteries

The Brief Test of Adult Cognition by Telephone (BTACT) was developed by Tun and Lachman (2006) in order to assess cognitive changes in normal aging. Specifically, the BTACT assesses for episodic verbal memory, working memory, executive functioning, and processing speed (i.e., word list recall immediate, backward digit span, category fluency, Stop and Go Switch Task (SGST), number series, the 30 second and Counting Task (30 – SACT), and word list recall delayed; Tun & Lachman, 2006). Specifically, the researchers compared adults across the life span by splitting them into three groups younger (< 40 years old), middle aged (40-59), and older (< 60 years old). An ANOVA was used to compare age groups in each domain that showed significant differences between groups for each of the domains which are as follows; immediate memory, F (2,81) = 8.40, $p < 0.001$; delayed memory, F (2,81) = 14.87, $p < 0.001$; working memory, F (2,79) = 3.37, $p = 0.039$; verbal fluency, F (2,78) = 5.23, $p = 0.007$; speed, F (2,81) = 13.84, $p < 0.001$ and reasoning, F (2,80) = 4.12, $p = 0.020$ (Tun & Lachman, 2006). Additionally, the researchers controlled for educations and found effects were significant for verbal fluency, F (1,77) = 6.93, $p = 0.010$; reasoning, F (1,79) = 9.04, $p = 0.004$; and working memory, F (1,78) = 7.35, $p = 0.008$ (Tun & Lachman, 2006). However, education effects was not significant for immediate, F (1,80) = 2.65, $p = 0.107$, or delayed memory, F (1,80) = 1.89, P = 0.173, or for speed, F (1,80) = 2.62, $p = 0.109$ (Tun & Lachman, 2006). A follow up study was conducted by Lachman et al. (2014) to determine the psychometric properties of the BTACT in comparison to an in-clinic evaluation. Two hundred and ninety-nine adults were administered both the BTACT and in-depth cognitive battery with ages ranging from 34-85 and a mean education of 15.36 (SD = 2.63; Lachman et al., 2014). The Boston cognitive battery was administered in person and took approximately 90 minutes; it included tests of forward digit span, backward digit span, serial sevens, verbal ability, letter series, and Raven's advanced progressive matrices, and digit symbol substitution (Lachman et al., 2014). Both the Boston cognitive battery and the BTACT were administered within two years of each other (Lachman et al., 2014). Lachman et al. (2014) ran comparison correlations in order to obtain concurrent validity between measures administered face to face and via telephone. All correlations between the BTACT test of backward digit span, category fluency, number series, 30 SACT with cognitive factors of short-term memory, verbal ability, reasoning, and processing speed were considered significant despite correlations being limited at best. Specifically, correlations between these tests ranged between .31 to .54 and were significant ($p < 0.001$; Lachman et al., 2014). Stronger correlations were noted with overall composite scores between BOLOS and BTACT with r (292) =.73, $p < 0.001$; Lachman et al., 2014). Although many of these correlations were significant, they are weak at best. Additionally, a large majority of Lachman et at. (2014) BTACT Test correlated with face-to-face administered tests, thereby questioning if BTACT individual tests are actually measuring their designated cognitive domain.

Attention and Working Memory

Digit Span. As stated above, a study conducted by Rankin et al. (2005) compared multiple neuropsychological assessments in both face-to-face administration and telephone administration for the AREDS populations. The study included 1,738 participants with a mean age of 74.9 (5.0) and compared the participants with a face-to-face battery that included MMSE, verbal fluency (letter fluency and animal fluency), Wechsler Memory Scaled-Revised (WMS-R) Logical Memory I and Logical Memory II, Buschke Selective Reminding Test, and Digits Backwards with a telephone battery that included all assessments as the in-clinic battery with the exception of the MMSE and the Buschke Selective Reminding Test (Rankin et al., 2005). The researchers compared both the face to face and telephone administration for an estimated correlation coefficient. Rankin et al. (2005) initially ran an unadjusted correlation and found a weak correlation between the in-clinic M = 6.4 (1.9) and telephone M = 7.1 (2.4) administrations. However, when the correlation analysis was adjusted for age and depression, it yielded a stronger correlation for Digit Span M = 6.4 (0.3) and telephone M = 7.1 (0.5) administrations ρ = 0.79 (95% CI, 0.76 - 0.81) thereby validating telephone administration for Digits Backward.

Another group of researchers, Bunker et al. (2017), administered both in person and telephone batteries to 50 participants who participated in the sub-study from Successful Aging after Elective Surgery (SAGES). Participants had a mean age of 74.9 (4.1), a mean education of 14.9 (2.5), were English speaking, and scheduled to undergo elective surgery with the anticipated length of stay of at least three days (Bunker et al., 2017). Exclusion criteria included evidence of dementia, active delirium or hospitalization within three months, legal blindness or severe deafness, history of schizophrenia, and/or history of alcohol abuse/withdrawal (Bunker et al., 2017). As part of the stages study, every six months following their elective surgery, subjects underwent neuropsychological test battery in person and for the present sub-study, a 30-minute telephone neuropsychological battery was administered to volunteer subgroup within 2-4 weeks of the in-person interview (Bunker et al., 2017). Bunker et al. (2017) included Hopkins Verbal Learning Test – Revised (HVLT-R), Digit Span Forwards and Backwards, Verbal and Semantic Fluency, and a modified version of the Boston Naming Test (BNT) short form; however, the researchers did not include Trails A and B, Visual Search and Attention test, and the RBANS Digit Symbol Substitution because they require pen and paper. Bunker et al. (2017) calculated differences in scores by assessment method by calculating mean differences in comparing using the paired t-test statistic and found no significant difference between Digit Span Forwards and Backwards with in person M = 17 (3.7) and telephone M = 19 (4.0) administration. However, there was a significant moderate correlation between administration methods for Digit Span Forwards and Backwards (r = .50, $p < 0.01$, 95% CI, 0.25, 0.68; Bunker et al., 2017). Given there was no significant difference between the means, this may provide understanding into normative information.

A similar study was conducted by Wilson et al. (2010) to assess differences between method of neuropsychological test administration on 1,584 older adults with a mean age of 71.1 (11.2) and a mean education of 14.2 (3.0) and approximately one third were administered the telephone battery. The test battery included Digit Span Forward, Backward, and Ordering, immediate and delayed recall from story A (WMS-R), and Semantic memory (fluency of Animals and Vegetables separate 1-min trails) all of which can be administrated in 10-15 minutes (Wilson et al., 2010). Wilson et al. (2010) spilt participants into two subgroups; dementia and no dementia and found that the dementia subgroup was older (79.2 vs 68.6, t [902] = 21.3, $p < 0.001$) with less education (13.2 vs 14.5, t[1,504] = 7.0, $p < 0.001$) when compared with the no dementia group. Wilson et al. (2010) ran a series of linear regression models with an indicator for telephone verse in person test administration while controlling for confounding effects of age, sex, and education. Additionally, Wilson at al. (2010) ran separate linear regression models for both dementia and no dementia. The researchers found for the working memory cognitive domain, which includes digit span forward, backward, and ordering, that mode of administration accounted for 1.4% of the variance $p < 0.001$ in those with no dementia. However, in those with dementia, the linear regression model was not significant thereby indicating no differences between mode of administration for the working memory cognitive domain in those with dementia period of note. Wilson et al. (2010) did not provide psychometric data for means and standard deviations for in person versus telephone assessment. These researchers also did not provide additional statistical information regarding their linear regression models including degrees of freedom and F change values that would increase readers' ability to better understand statistical analysis.

Another study conducted by Rapp et al. (2012) assessed modes of neuropsychological test administration specifically telephone verse face to face administration in 95 nondemented women who were divided among four groups; face/face, face/telephone, telephone/face, and telephone/telephone which were administered approximately six months apart. The neuropsychological test battery was developed in order to assess attention, concentration, verbal learning and memory, verbal fluency, working memory, and executive functioning. Specifically, the test included the California verbal learning test, letter fluency and category fluency (F, A, S and Animals), and the Digit Span-Forward and Backward from WMS-II (Rapp et al., 2012). Rapp et al. (2012) assessed test-retest reliability with Pearson correlation coefficients for each administration by the same mode in the six-month interval. Concurrent validity was assessed by a fixed effect general lineal models for data collected from two time periods for both modes of administration in order to assess the telephone batteries ability to detect changes. Additionally, they examined cross sectional means for each test and mean changes over time. Rapp et al. (2012) found no significant differences between mean scores for face to face and telephone administration at baseline for both Digit Span Forward and Backward ($p = 0.88$ and $p = 0.44$ respectively). Additionally, Rapp et al. (2012) compared estimates of relative bias between face to face and telephone administrations and found no statistically significant bias for Digit Span Forward (M = -

0.01, SE = 0.11, *p* = 0.94) and Digit Span Backward (M = 0.28, SE = 0.12, *p* = 0.02). Additionally, when comparing performance of Digit Span Forward and Backward between non-Whites and non-Hispanic Whites with administration mode, Rapp et al. (2012) found no significant differences. Rapp et al. (2012) assessed change in standard deviation and mean scores for both modes of administration and found no differences, thus indicated that for older adult women with normal to mildly impaired cognition that Digit Span Forward and Backward is a reliable and valid test to administer over the telephone.

Language

As noted above, Bunker et al. (2017) administered both in-person and telephone batteries to 50 participants who participated in the sub-study from SAGES. These researchers modified the Boston Naming Test (BNT) short version with 15 items to assess auditory naming that uses vocabulary and confrontation naming. Specifically, the interviewer read a short sentence describing an object and the participant was asked to name it (Bunker et al., 2017). The interviewer was allowed to provide a phonemic queue and if the participant was able to answer correctly with the phonemic, only a half point was awarded (Bunker et al., 2017). The list of objects to name in the telephone interview was the same as the in-person interview and was in the same order that the objects were initially presented (Bunker et al., 2017). Bunker et al. (2017) compared mean differences in scores for each mode of administration by using the paired t-test statistic and assessed the agreement of mode of administration test scores estimated by the Pearson correlation coefficient. When comparing mean scores for face to face (M = 14 [1.7]) and telephone administration (M = 14 [1.6]), participants scored lower with the telephone administration with a mean difference of -0.26 (95% Cl -0.52, -0.01; Bunker et al., 2017). Additionally, Bunker et al. (2017) found a strong correlation for the agreement between modes of administration (r = 0.85, 95% CI; 0.75, 0.91, $p < 0.01$) thus indicting agreement between modes of administration. Although there is agreement between administration, this does not always indicate equivalents; as such, this still leaves us with the question on which norms will be best or if developing new norms is best.

Executive Function

COWAT. As previously documented, a study conducted by Rankin et al. (2005) compared multiple neuropsychological assessments to compare mode of administration between telephone and face to face. Rankin et al. (2005) administered both Verbal Fluency (F, A, S) and Category Fluency (Animals) with no modifications. Rankin et al. (2005) compared mode of administration with a correlation analysis for raw scores and predicted scores from a regression analysis that was adjusted for both age and depression. Rankin et al. (2005) initially ran an unadjusted correlation and found a moderate but significant correlation for both Verbal Fluency ($\rho = 0.79$, 95% CI, 0.76-0.81) and Category Fluency ($\rho = 0.62$, 95% CI, 0.68, 0.65) with both modes of administration. When the analysis was adjusted for age and depression, it yielded a similar correlation for Verbal Fluency face to face M = 38.9 (13.3) and telephone M = 37.8 (14.0) administrations $\rho = 0.71$ (95% CI, 0.68, 0.74; Rankin et al., 2005). Rankin et al. (2005) had a similar finding for Category Fluency face to face M = 17.6(4.9) and telephone M = 16.6 (5.0) administrations $\rho = 0.82$ (95% CI, 0.81, 0.84). This confirms a significant linear association between face to face and telephone adjusted scores implying that letter fluency and animal category fluency instruments give constant scores either through telephone administration or an in-person administration when adjusting for age and depression at the time of administration. Although there is a significant linear association, these researchers did not run mean comparisons making it difficult to validate current normative data for telephone assessment.

As previously stated, Bunker et al. (2017) administered both in person and telephone batteries to 50 participants who participated in the sub study from SAGES to compare mean differences in scores for each mode of administration by using the paired t-test statistic and assessed the agreement of mode of administration; test scores were estimated by the Pearson correlation coefficient. Bunker et al. (2017) administered both Category fluency (grocery store items) and Phonemic fluency (F, A, S) and no modifications were made. In regard to comparing means for Phonemic Fluency for face to face (M = 45 [13.8]) and telephone administration (M = 44 [14.6]), participants scored lower with the telephone administration with a mean difference of -1.40 (95% Cl -3.05, 0.25; Bunker et al., 2017). There was a strong correlation for the agreement between modes of administration for Phonemic Fluency (r = 0.92, 95% CI; 0.86, 0.95, $p<0.01$; Bunker et al., 2017). However, with Category Fluency, participants had higher scores with telephone administration M = 25 (6.3) when compared with face-to-face administration M = 24 (5.9) with a mean difference of 1.12 (95% CI; -0.36, 2.60; Bunker et al., 2017). Although the Pearson correlation was statistically significant for Category Fluency, it is considered moderate at best (r = 0.63, 95% CI; 0.43, 0.77, $p<0.01$; Bunker et al., 2017), meaning that mode of administration may have impacted participant's ability to take the test properly.

As mentioned above, a study conducted by Wilson et al. (2010) was completed to assess differences between mode of neuropsychological test administration on 1,584. Semantic memory (fluency of Animals and Vegetables separate 1-min trails) were administered both face to face and telephone (Wilson et al., 2010). Wilson et al. (2010) spilt participants into two subgroups: dementia and no dementia. Wilson et al. (2010) ran a series of linear regression models with an indicator for telephone verse face to face test administration while controlling for confounding effects of age, sex, and education and ran separate linear regression models for both dementia and no dementia. A factor analysis found that semantic fluency loaded on two possible factors, either semantic or declarative memory. Unlike Digit Span, mode of administration did not account for a significant amount of the variance for semantic memory in both the dementia (< .1%) and no dementia (< .1%; Wilson et al., 2010).

As reported earlier, a study conducted by Rapp et al. (2012) was done to assess modes of neuropsychological test administration specifically telephone verse face to face administration in 95 nondemented women who were divided among four groups. Letter Fluency and Category Fluency was administered using F, A, S and Animal prompts with no modifications. Rapp et al. (2012) assessed test-retest reliability, cross sectional means for each test, and mean changes over time. Rapp et al. (2012) found no significant differences between mean scores for face to face and telephone administration at baseline for both Letter Fluency and Category Fluency ($p = 0.43$ and $p = 0.14$, respectively). Additionally, Rapp et al. (2012) compared estimates of relative mean bias between face to face and telephone administrations and found no statistically significant bias for Verbal Fluency (M = -0.09, SE = 0.08, $p = 0.26$) and Category Fluency (M = -0,08, SE = 0.10, p

= 0.2). Additionally, when comparing performance between non-Whites and non-Hispanic Whites with administration mode, Rapp et al. (2012) found no significant differences. Rapp et al. (2012) assessed change in standard deviation and mean scores for both modes of administration and found no differences thus indicating that for older adult women with normal to mildly impaired cognition, Verbal, and Category Fluency is a reliable and valid test to administer over the telephone.

Memory

Logical Memory. As previously stated, Rankin et al.'s (2005) study compared multiple neuropsychological assessments to compare mode of administration between telephone and face to face. Rankin et al. (2005) administered WMS-R Logical Memory I and II with no modifications. A comparison for mode of administration was done with a correlation analysis for raw scores and predicted scores from a regression analysis that was adjusted for both age and depression. Rankin et al. (2005) reported Logical Memory I face to face administration had a mean of 38.0 (10.6), telephone administration had a mean of 42.6 (11.2), Logical Memory II face to face administration had a mean of 22.2(8.3), and telephone administration had a mean of 25.4(9.1). Similar to the other results, Rankin et al. (2005) found a weak to moderate but significant correlation for unadjusted scores for both Logical Memory I ($\rho = 0.67$, 95% CI, 0.64, 0.69) and Logical Memory II ($\rho = 0.71$, 95% CI, 0.68, 0.7; Rankin et al., 2005). When the scores were adjusted for age and depression, Rankin et al. (2005) found a stronger correlation for both Logical Memory I ($\rho = 0.87$, 95% CI, 0.86, 0.88) and Logical Memory II ($\rho = 0.86$, 95% CI, 0.84, 0.87). Again, this confirms a significant linear association between both face to face and telephone adjusted and unadjusted scores as such logical memory gives consistent scores across both modes of administration. Although the logical memory is considered correlated across modes of administration, no information was provided by these researchers and no proper normative data to use given the no mean and standard deviation differences between both telephone and face to face administration.

Another study conducted by Wilson et al. (2010) was completed to assess differences between mode of neuropsychological test administration on 1,584 individuals. Semantic memory (fluency of Animals and Vegetables separate 1-min trails) were administered both face to face and telephone (Wilson et al., 2010). Wilson et al. (2010) split participants into two subgroups: dementia and no dementia. Wilson et al. (2010) ran a series of linear regression models with an indicator for telephone versus face-to-face test administration while controlling for confounding effects of age, sex, and education and ran separate linear regression models for both dementia and no dementia. Wilson et al. (2010) only administered Story A from Logical Memory I and II WMS-R with no modifications. Wilson et al. (2010) ran a factor analysis and found that story A loaded on two possible factors either episodic or declarative memory. In regards to the impact of mode of administration, linear regression models indicated that administration method was not significant as it accounted for < 0.1% of the variance for both dementia and no dementia groups (Wilson et al., 2010).

CVLT. As stated earlier, a study conducted by Rapp et al. (2012) compared adult women to assess modes of neuropsychological test administration who were divided among four groups. Rapp et al. (2012) administered the modified versions of the CVLT as only three of the five immediate recall lists are given for a total score of 48; however, the rest of the task remained intact. Rapp et al. (2012) assessed test-retest reliability, cross sectional means for each test, and mean changes over time. Rapp et al. (2012) found no significant differences between mean scores for face to face and telephone administration at baseline for all possible scores on the CVLT (Recall List A, Recall List B, Short Delay Free Recall, Long Delay Free Recall, Short Delay Cued Recall, Long delay Cued Recall, and Recognition) as all *p* values were above 0.43. Additionally, Rapp et al. (2012) compared estimates of relative mean bias between face to face and telephone administrations and reported no statistically significant bias for any of the CVLT scores below $p < 0.01$. However, when looking at the numbers provided Short Delay Free Recall (M = 0.02, SE = 0.10, $p = 0.04$) and Recall list B (M = 0.24, SE = 0.11, $p = 0.3$) do fall below the significance level of $p < 0.05$. Additionally, when comparing performance between non-Whites and non-Hispanic Whites with administration mode, Rapp et al. (2012) found that on the Recognition Subtest, Non-Whites showed worse performance on telephone administration ($p = 0.0002$). The change in standard deviation and mean scores for both modes of administration showed no differences, thus indicating that for adult women with normal to mildly impaired cognition, the CVLT is a reliable and valid test to administer over the telephone.

HVLT-R. As reported earlier, Bunker et al. (2017) compared both in person and telephone neuropsychological test batteries with 50 participants who were involved in the sub study from SAGES to compare mean differences in scores for each mode of administration by using the paired t-test statistic and assess the agreement of mode of administration test scores that is estimated by the Pearson correlation coefficient. Bunker et al. (2017) administered the HVLT-R with no modifications and reported scores for HVLT-R Total Recall, HVLT-R Delayed Recall, HVLT-R Discrimination Index, and

HVLT-R Retention Percentage. When comparing means of difference for modes of administration, Bunker et al. (2017) found that the largest mean difference was with Total Recall. On Total Recall, Delayed Recall and Discrimination Index; participants all scored higher on the telephone administration. The Pearson correlations for tests of Total Recall, Delayed Recall, and Discrimination Index were statistically significant at the $p < 0.01$ level. However, the correlation for Retention Percentage scores for mode of administration was not statistically significant; participants scored higher with face-to-face administration. The researchers reported limited concern regarding these findings.

Table 1

In-person Versus Telephone Means, Standard Deviations, Correlations, and Paired Tests for the HVLT-R

Test	Face to face M(SD)	Telephone M(SD)	Correlation (95% CI)	Mean Difference (95% CI)
Total Recall	27 (5.8)	28 (5.6)	0.87 (0.79, 0.93) *	1.64 (0.82, 2.46)
Delayed Recall	9 (2.5)	10 (2.2)	0.75 (0.60. 0.85) *	0.28 (-0.20, 0.76)
Discrimination Index	10 (1.4)	10 (1.3)	0.62 (0.41, 0.77) *	0.30 (-0.04, 0.64)
Retention Percentage			0.27 (-0.01, 0.51)	-1.37 (-6.15, 3.40)

*Significant at the $p < 0.05$

Eligibility Criteria

What Does the Neuropsychologist Need?

Unlike face-to-face neuropsychological assessment, telephone neuropsychological assessment needs limited materials. Despite the limited materials needed, there are some things for the neuropsychologist to consider prior to undertaking telephone assessments. Many of the above outlined studies reported needing limited technology but did not discuss the phone systems used. As such, it may be assumed they used typical landlines or mobile phones. Additionally, many studies discussed the need to have caregiver assistance. Monteiro et al. (1998) asked for caregiver assistance to help assess successful completion of three step commands and ability to write a sentence. Furthermore, requests were made for participants to be in a quiet room where they were free of distractions and no orientating information was available. Much of the trust is placed upon the neuropsychologist to believe that the patient is not cheating. As such, it will be important to vet the patients during initial visits in order to ensure they will not write down word lists and or use orientating information during assessments.

What Populations are Best suited for Telephone Assessments?

Demographics. Many of the above-mentioned studies conducted research with specific populations. One study only used female participants (Rapp et al., 2012) whereas the vast majority of the studies had higher participation with female participants (Bunker et al., 2017; Monteiro et al., 1998; Wilson et al., 2010; Wong et al., 2015). In regard to age, there are always a wide range of participants in the studies. Specifically, most often with the cognitive screeners research worked with older adults typically above 65. The BTACT had the widest range of ages used, with an age range of 34 – 85 and a (mean age of 58[13];Lachman et al., 2014). In the research that was conducted to compare neuropsychological batteries, participants were typically older adults. Wilson et al. (2010) reported having participants ranging in age from 28 to 99; however, they reported a mean age of 71.1 (11.2). Across all studies, there were limited participants who identified as non-White. Only one study reported cognitive differences based on ethnicity (Rapp et al., 2012). Additionally, the research conducted by Wong et al. (2015), was conducted in Hong Kong and administered in Cantonese; as such, the research regarding the MoCA 5-minute protocol may not generalize to another population. Consistently, across all research studies, there was a higher level of education typically with a mean education of 14 years.

Cognition. In the studies that were evaluating cognitive screeners, the researchers often compared normal cognition with the cognitively impaired. Cognitive severity ranged from normal cognition to mild impairment to AD or dementia (Brandt et al., 1988; Fong et al., 2009; Monteiro et al., 1998). Specifically, Wong et al. (2015) researched those with stroke or TIA by comparing normal cognition and cognitively impaired. Wong et al. (2015) also found success with the MoCA 5-minute protocol to help with differentiating cognitive impairment and cognitive impairment with executive functioning. In the studies outlined above, there was varying participation from older adults with cognitive difficulties. Only one study was conducted with older adults who were described as having no dementia or dementia (Wilson et al., 2010). The other studies reported only including participants who were non-demented or generally healthy (Bunker et al., 2017; Rankin et al., 2005; Rapp et al., 2012). Additionally, in regard to the Lachman et al. (2014) BTACT study, the researchers reported that the participants indicated their cognitive functioning in health as generally healthy and the researchers did not indicate using cognition as an exclusion criterion.

Exclusion Criteria. Most of the above-mentioned studies discussed level of hearing in their discussions period; however, few did use it as an exclusion criterion. Specifically, the studies which undertook comparisons for neuropsychological batteries typically indicated using poor hearing as an exclusion criterion (Bunker et al., 2017; Rankin et al., 2005; Rapp et al., 2012; Wilson et al., 2010). Bunker et al. (2017) provided the following as exclusion criteria evidence of dementia: active delirium or hospitalization within three months, legal blindness or severe deafness, history of schizophrenia, and/or history of alcohol abuse/withdrawal. Most studies did have participants answer questionnaires regarding mood symptoms; however, depression often was held constant in running comparison studies because depression can impact cognitive functioning.

Implications for Clinical Practice

Benefits of Telephone Assessments

Research would not be conducted for telephone assessments if there were not the potential for benefits for these assessments. Through most of these studies, the purpose of the research was to assess the feasibility of telephone assessments to increase accessibility for patients and participants in research studies. Additionally, many of the researchers were able to develop additional cutoff scores to use for telephone assessments when using cognitive screeners such as the MMSE, TICS, and MoCA. Although not assessed directly, a few researchers notice that patients reported being more at ease during the telephone administration than during in person administrations (Fong et al., 2009; Monterio et al., 1998). Additionally, many of the research studies indicated that the telephone administration was shorter to administer than the face-to-face administration. Shorter administration time may increase compliance and patient willingness to complete neuropsychological assessments because traditional neuropsychological assessments are typically a few hours long.

Disadvantages of Telephone Assessments

Although there are many advantages to use telephone neuropsychological assessments, there are some disadvantages. The study conducted by Bunker et al. (2017) discussed some of these disadvantages around hearing and potential loss of control during testing. Specifically, there were at times issues with the ability to verify the participant was in a quiet and calm environment and the able to ensure compliance with not writing down the word list or digits provided. In one of the studies that assessed for a cognitive Screener, TICS, found that participants in the telephone trial often showed higher orientation scores (Fong et al., 2009). The researchers noted there may be increased distractions that could have contributed to lower test scores on some assessments during the telephone trials (Bunker et al., 2017). Additionally, often tests were modified for the telephone administration specifically, modifications were made to the BNT, CVLT, MoCA, and MMSE. Furthermore, many of the research studies had limited sample sizes; as such, these limited sample sizes are likely not a complete representation of the general population for example. Bunker et al. (2017) had a sample size of 50 and Rapp et al. (2012) had a sample size of 110. Additionally, there were limitations in the telephone neuropsychological batteries administered. All research studies left out practice exam, visual spatial abilities, and visual processing speed.

Missing Pieces

Many of these studies indicated that they were for research purposes to help address difficulty with follow-up and seeing participants who were farther away from research sites; as such, these studies did not address needs associated with clinical neuropsychologist. Many of the studies assessing the feasibility of cognitive screeners for telephone assessment were able to derive additional cutoff scores for modified of the TICS, MoCA, and MMSE. Additionally, these studies were able to run specificity and sensitivity for their ability to accurately identify healthy older adults without cognitive difficulties and those with cognitive impairments (Brandt et al., 1988). Brandt et al. (1988) reported the ability for the 5-minute MoCA to differentiate between normal cognition and cognitive impairment specifically with executive dysfunction. Additionally, because many of the neuropsychological battery research only assessed those without cognitive impairment, it does not help the clinical neuropsychologist recognize the telephone battery's ability to differentiate between normal cognition and impaired cognition. Additionally, no information was provided in the research on how to address report writing in clinical settings because the mean and standard deviation difference can impact normative data and in turn impact the tests ability to accurately assess impairment. Lastly, future research needs to address time elapsed between testing because over time, patients and participants can experience change in cognition that may impact their test scores. This addresses the needs of the neuropsychologist when report writing and scoring. Rankin et al. (2005) and Rapp et al. (2012) both reported significant time elapsed between testing that may impact the ability to accurately assess mode of assessment due to potential changes in cognitive in patients spanning 4-12 months.

Emerging Technology

Smart Phone Applications

Given the development of smartphones and applications for smartphones, many researchers are currently studying the feasibility and reliability of using smartphone applications to administer cognitive assessments to assess cognitive functioning in the older adult community. Brouillette et al. (2013) conducted a study regarding the development of a smartphone-based application to measure cognitive function in the older adult population. The color shaped hearts is a test of cognitive processing speed where the participants are asked to correctly match as many shapes with their current responding color as quickly as possible. Specifically, it consists of paired colors and shapes and these colors/shapes appear on the top of the screen and serve as a legend. At the bottom of the screen are colored blocks that correspond to colors in the legend. Participants are to use the pads to respond to coordinate the colors with a shape that appears on the screen. They are given approximately 30 seconds to respond. The test records the number of attempts and number of correct attempts over a two-minute testing interval (Brouillette et al. 2013). Brouillette et al. (2013) conducted a study using 57 community dwelling adults with a mean age of 67 and mean education of 16 years. They were considered healthy older adults because they did not have a diagnosis of dementia or other neurological condition **(**Brouillette et al. 2013). The researchers compared the color shape test with typical neuropsychological battery that included MMSE, Digits Forwards and Backwards WMS-R, Digit Symbol Test WIAS-R, Trail Making Test Part A and B, Verbal Fluency (Animal and Vegetable), Logical memory I and II WMS-R, and BNT 30 odd items. Brouillette et al. (2013) found convergent validity for multiple measures including Digit Span, Trail Making Tests, and Digit Symbol test ($r = 0.427$, $p < 0.0001$; $r = -0.651$, $p < 0.0001$; $r = 0.508$, $p < 0.0001$, respectively). The color shape test was also correlated with the MMSE ($r = 0.515$, $p < 0.001$; Brouillette et al., 2013).

Moore et al. (2017) completed a systematic review on current mobile cognitive assessments and included 12 articles that broke down to eight studies conducted in

European countries, four of which were conducted in the United States. Specifically, the majority assessed community dwelling healthy older adults and only four studies examined adults with illnesses (Moore et al., 2017). The review found that use of smart devices is generally feasible among research participants and reported good psychometric properties for self-administered cognitive assessments. Takeaways include mobile cognitive assessments help with enhancing the sensitivity of assessing slight cognitive changes while someone is in their home environment; found to be more sensitive to have a screening tool for diagnosing early cognitive decline; provide the ability to assess cognitive difficulties over time including initial baseline and continuous assessment of cognitive data over the course of the treatment. Allowing for sensitive assessments of cognitive change that may occur due to age-related decline, neurological diseases, and or psychiatric illnesses allows for the ability to assess between and within day variability of cognition that will help with examining sensitivity of side effects to treatments, understanding confusion, and delirium (Moore et al., 2017).

Clinical Pearls

Table 2 provides a breakdown of the research reviewed based on assessment administered and if new normative data needs to be developed. Additionally, information is provided if modifications were made to the assessment. Table 2 can be used as a quick reference guide for clinicians when deciding what assessments to utilize during remote computerized assessments.

Table 2

Does Mode of Administration Matter for Telephone Assessments

Test	Valid	Cutoff score	Modifications	New normative data needed
TICS	Yes	31	Yes	
TICS-M	Yes		Yes	
TICS-30	Yes	25-30	Yes	
TICS-40	Yes	32-40	Yes	
T-MoCA	Yes	18-19	Yes	
T-MoCA Short		10-11	Yes	
BTACT	No	N/A	Yes	Weak correlations
Digit Span	Yes	N/A	No	Yes, more research needs done to gain equivalence for both means and standard deviations
BNT	Yes	N/A	Yes	Yes, one study found mean differences; as such, more research needs to be done to gain equivalence for both means and standard deviations
COWAT	Yes	N/A	No	Yes, mean differences were found; however, more research needs to be done to determine if there are standard deviation differences
Logical Memory	Yes	N/A	No	Yes, no information was provided regarding mean or standard deviation differences; as such, further research needs to be conducted
CVLT	Yes	N/A	Yes	There were mean differences based on race for mode of administration
HVTL-R	Yes	N/A	No	Mean differences were found; however, no information was provided regarding standard deviation differences

*N/A = not applicable

CHAPTER III: TELEHEALTH: COMPUTERIZED NEUROPSYCHOLOGICAL ASSESSMENTS

As technology advances, neuropsychological assessments advance with it; as such, the following chapter will compare modes of administration for computerized neuropsychological assessments versus traditional paper pencil assessments. Specifically, this chapter will review computerized assessment that have been developed from traditional paper pencil assessments. Although many computerized assessments that have been developed, the purpose of this literature review is to determine if mode of administration impacts original normative data provided for paper and pencil traditional and face to face in neuropsychological assessments. This chapter will review current research that compares computerized neuropsychological assessments with traditional paper pencil neuropsychological assessments to help determine if mode of administration does impact normative data. Additionally, this chapter will provide neuropsychologists pertinent information on populations best suited for computerized neuropsychological assessments, technology needed, and considerations for both diagnosis and report writing.

Overview of Current Research

Cognitive Screeners

A study conducted by Saxton et al. (2009) was completed to compare the sensitivity and specificity of the Computer Assessment of Mild Cognitive Impairment (CAMCI) with the MMSE to identify mild cognitive impairment in a population of 524 order adults who did not have dementia. The CAMCI was developed specifically for older adults who may be uncomfortable with computers; as such, it has a simple design and runs on a tablet computer. The CAMCI used modified standard neuropsychological tests of attention, executive functioning, working memory, and variable in visual memory (Saxton et al., 2009). Specifically, the modified paper pencil tasks include star task, forward digits span, word recognition, word recall, picture recognition, go no go test, digit reverse span (Saxton et al., 2009). Additionally, a second part of the test uses virtual reality in which the individual moves through a grocery store on a shopping trip which is intended to resemble everyday experiences (Saxton et al., 2009). Specifically, the shopping trip is where the participants are asked to navigate a virtual world and as they are on their way, they are told they must run several errands in addition to the shopping trip; this allows for a potentially more ecologically valid test as it includes recognition memory, incidental recall, and perspective memory (Saxton et al., 2009). The sample included 296 participants who were identified as having normal cognition and 228 as being in the range of MCI (Saxton et al. 2009). Saxton et al. (2009) found that the CAMCI had a better sensitivity and specificity than MMSE as its sensitivity was 86% and specificity was 94% whereas when using a cutoff score of 28 on the MMSE, sensitivity was 45% and specificity was 80%.

A study by Dion et al. (2020) examined cognitive constructs of the digital clock draw and compares MCI and Non-MCI non-demented older adults' performances. The digital Clock Draw Test (dCDT) has participants draw a clock and copy a clock with the use of a digital pen that utilizes software for scoring and graphomotor speed (Dion et al., 2020). The dCDT obtains the following scores: Total Completion Time (TCT) – total time taken to draw the clock, Pre-Frist Had latency (PFHL) – time taken between drawing the first clock hand and the previous stroke, Post-Clock Latency (PCFL) time between completing the clock face and the first number, Clock Face Area (CFA) – circumference of the circle, and “Think” versus Ink time (Dion et al., 2020). Additionally, Dion et al. (2020) compared these with corresponding cognitive domains with traditional neuropsychological assessments: processing speed – Digit span, Stroop color word and reading conditions, TMT A, working memory – letter number sequencing, DS backward, Spatial span, language – BNT, COWA (animal), and declarative memory – Logical Memory I and II, HVLT-R. Dion et al. (2020) ran correlations between dCDT variables with cognitive domains while controlling for age and cognitive reserve. Total Completion Time (TCT) was associated with slower performance on processing speed test ($r = -0.284$, $p < 0.001$) and worse performance on working memory ($r = -0.240$, $p = 0.001$; Dion et al., 2020). Additionally, the TCT was also significantly associated with a negative correlation with language and declarative memory in the command condition (Dion et al., 2020). Pre-Frist hand latency (PFHL) was initially negatively correlated with working memory; however, the effect sizes were small, and the correlation was no longer present after correcting for multiple comparisons (Dion et al., 2020). Post-Clock Latency (PCFL) was initially negatively correlated with processing speed; however, after correcting for multiple comparisons, the correlation was no longer present. No relationship was noted with Clock Face Area (CFA; Dion et al., 2020). In the command condition, the univariate analysis comparing MCI status found a significant difference with TCT in the MCI group when you had a slower TCT (Dion et al., 2020). Overall, TCT had the strongest relationship to traditional neuropsychological testing performance including processing speed, working memory, language, and declarative memory. This is consistent with previous research on traditional versions of Clock draw.

Attention and Working Memory

Digit Span. Vermeent et al. (2020) evaluated a digital version of a traditional neuropsychological battery to determine if the digital version has the same factor loadings as would be expected with the traditional paper pencil tasks. Vermeent et al. (2020) administered both digit span forward and backward with the use of an iPad where the numbers were automatedly presented to the participant and the examiner recorded the answers. Digit span forward and backward were scored using the iPad software. Vermeent et al. (2020) found that digit span loaded on working memory through the use of the neuropsychological consensus model ($z = 8.31$, $p > 0.001$ and $z = 8.95$, $p > 0.001$) thus indicating that the digital version of digit span forward and backward measures the same cognitive domain as the paper pencil version. Spreij et al. (2020) administered the same digital neuropsychological battery (d-NPA) as Vermeent et al. (2020) through Phillips Research. These researchers sought to assess the feasibility and accuracy traditional norms for the d-NPA in those with an acquired brain injury (Spreij et al., 2020). In order to assess if traditional norms are applicable to computerized testing, they expected that less than ten percent of the healthy controls would perform below the 10^{th} percentile based on Lezak's distribution. When the analysis was conducted, Spreij et al. (2020) found that stroke (16.1%) and TBI (37.7%) participants had higher percentages of abnormal performance on Digit Span and, as to be expected, only 8.8% of healthy controls had an abnormal performance. This indicated that traditional paper pencil norms for Digit Span are applicable for the tablet version of Digit Span.

Processing Speed

Trail Making Test A. As mentioned above, Vermeent et al. (2020) evaluated a digital neuropsychological battery. The digital version of Trail Making Test A (TMT A) was administered through the use of an iPad where the patient connects numbers 1 to 25 as fast as they can and is automatically scored on the iPad. Vermeent et al. (2020) found that TMT A loads on processing speed the use of the neuropsychological consensus model ($z = 9.39$, $p > 0.001$). A similar study conducted by Spreij et al. (2020) used the same version of the d-NPA research battery as Vermeent et al. (2020) to conduct an analysis regarding if traditional paper pencil norms are equivalent and/or applicable to the tablet version of TMT A. When the analysis was conducted, Spreij et al. (2020) found that 42.9% of the stroke participants and 40% of the TBI participants had an abnormal performance. However, with the healthy controls, 24.5% of the participants had an abnormal performance, which is more than should be expected given Lezak's distribution (Spreij et al., 2020). As such, traditional paper pencil norms for TMT A are not considered equivalent or acceptable for the tablet version.

Bracken et al. (2018) assessed the TMT adapted for the iPad by Parker-O'Brien to assess reliability and validity. The TMT for the iPad was administered using an iPad Air with the use of a stylus. Both modes of administration utilized traditional instructions, and errors were immediately corrected and marked on both paper pencil and iPad versions. Bracken et al. (2018) assessed test-retest reliability using both Pearson correlation and interclass correlations and assessed concurrent validity. Bracken et al. (2018) assessed 77 participants who were split into four groups to counterbalance order of administration. In regards to TMT A, test retest reliability was variable as only one group had an adequate Pearson R correlation (*r* (22) = 0.71; Bracken et al., 2018). Additionally, when comparing mode of administration, Bracken et al. (2018) did not find significance between the iPad version and the traditional paper pencil version. Of note, another analysis was conducted to examine impacts of handedness on performance. On TMT A, left handers performed slower on the iPad version (Bracken et al., 2018). This difference in handedness further proves additional norms will be needed for iPad versions of TMT.

Stroop. Vermeent et al. (2020) also administered a digital version of the Stroop task to evaluate if it had similar cognitive loadings as the paper pencil task. Stroop Color Naming and Interference was administered through the iPad where color names are presented and the clients are asked to name the color as quickly as they can or color names with incongruent color; however, scoring is the same as it is with the paper pencil version. Similar to previous results, Vermeent et al. (2020) found that Stroop Color Naming loaded on processing speed (z = 8.29, $p > 0.001$) and Stroop Interference loading on executive functioning (z = 9.21, $p > 0.001$).

A study conducted by Edwards et al. (1996) examined the effect of condition for the Stroop task with 27 young adults with a mean age of 21.4 using a between subjects design. Edwards et al. (1996) found a significant main effect for condition with comparing computer versus traditional task with how long it took participants to complete each subtest; for neutral word (F [3,75] = 3.34, $p < 0.05$) and color-word (F [3,75] = 7.02, $p < 0.001$). Participants tended to be faster on the computerized version on both subtests thus indicating that the card and computer versions are not equivalent.

Additionally, computer and manual versions may not be similar regarding norms given this finding and should include separate norms based on mode of administration.

Cancellation Test. Vermeent et al. (2020) administered both the Star-Cancellation Test (SCT) and the O-Cancellation Test (OCT) both of which requires the participant to cross out target stimuli on the iPad screen with distracting stimuli. Unlike paper pencil task, the digital task has automatic scoring, and all drawing is done through the use of an iPad (Vermeent et al., 2020). When analyzing for the factor loading using the neuropsychological consensus model, Vermeent et al. (2020) found that both cancellation tests loaded on the processing speed factor (SCT $z = 6.35$, $p > 0.001$ and OCT $z = 4.63$, $p > 0.001$). A similar study conducted by Spreij et al. (2020) used the same versions of SCT and OCT in the research battery of Vermeent et al. (2020) to analyze if traditional paper pencil norms are equivalent or applicable for the tablet version of SCT and OCT. When the analysis was conducted, Spreij et al. (2020) found that stroke (5.4%) and no TBI participants had abnormal performance on OCT; to be expected, only 3.8% of healthy controls had an abnormal performance. A similar performance was seen on SCT in that 1.8% of stroke participants and 6.7% of TBI participants had an abnormal performance; only 6.9% of healthy controls had an abnormal performance thus indicating that traditional paper pencil norms for both SCT and OCT are applicable for the tablet version.

Visuospatial Ability

Rey-O. Vermeent et al. (2020) administered the Rey-Osterrieth Complex Figure Test (ROCFT) copy as part of a larger digital battery to examine if the digital version of paper pencil tasks load on the same cognitive factors using the neuropsychological consensus model. The ROCFT copy was administered using an iPad where the participants were asked to copy a figure and all drawing was done on the iPad; however, scoring was the same as paper pencil tasks (Vermeent et al., 2020). The ROCFT loaded on the visual-spatial processing factor (z = 21.86, $p < 0.001$), which was to be expected. Spreij et al. (2020) used the same battery as Vermeent et al. (2020) to conduct an analysis to examine if traditional paper pencil norms are equivalent or applicable for the tablet version of ROCFT. When the analysis was conducted, Spreij et al. (2020) found that 30.4% of stroke participants and 34.4% of TBI participants had an abnormal performance. However, 16.4% of the healthy controls had an abnormal performance which is greater than 10%; this is to be expected based on Lezak's distribution (Spreij et al., 2020). Although the ROCFT copy loads on the visual spatial processing factor, it may be pertinent to provide separate norms for the tablet version.

Cube Drawing. Additionally, Spreij et al. (2020) administered cube drawing as part of their d-NPA to determine if traditional paper pencil normative data is acceptable for tablet versions. Cube drawing was administered on a tablet and was recorded automatically; however, scoring was still done by the neuropsychologist (Spreij et al., 2020). When the analysis was conducted, Spreij et al. (2020) found that 26.8% of stroke participants and 31.1% of TBI participants had an abnormal performance. However, 22.6% of the healthy controls had an abnormal performance which is greater than 10% based on Lezak's distribution (Spreij et al., 2020). As such, the traditional paper pencil norms for cube drawing are not applicable for the tablet version and may impact a neuropsychologist's ability to accurately gauge impairment.

Line Orientation. Askar et al. (2012) assessed 77 healthy volunteer undergraduates on The Line Orientation test across mode of administration. Askar et al. (2012) used the paper version from H developed from Benton et al. (1978). Askar et al. (2012) reported that the Line Orientation computerized test provides instructions that need to be read and automated scoring. All participants were administered both modes and administration that was approximately 22 days apart to reduce learning effects. Specifically, a correlation analysis was run to determine if both modes of administration were correlated, and *t*-tests were used to analyze mean differences. Total score correlation was significant for mode of administration ($r = .61$, $p < .05$; Askar et al., 2012). Of note, Askar et al. (2012) found a significant difference for mode of administration, $t(66) = 6.17$, $p < .05$, as the paper pencil version (M = 22.76, SD = 4.31) had higher scores than the computer version (M = 19.58, SD = 4.93). This indicates that the two versions are not equivocal and new normative data should be developed.

Executive Functioning

Trail Making Test B. Vermeent et al. (2020) administered Trail Making Test (TMT) B as part of a digital neuropsychological battery to analyze factor loadings for a digital test to see if they compare to the same loadings as paper pencil tasks. TMT B was administered using the iPad with automated scoring. TMT B loaded on the Executive Functioning factor using the Neuropsychological Consensus model ($z = 21.86$, $p < 0.001$) (Vermeent et al., 2020). Spreij et al. (2020) used the same d-NPA as Vermeent et al. (2020) to assess if TMT B norms were applicable or equivalent when comparing mode of administration. Specifically, 19.6% of stroke participants and 26.7% of TBI participants had an abnormal performance which was to be expected (Spreij et al., 2020).

Additionally, 3.1% of the healthy controls had an abnormal performance which was to be expected when using Lezak's distribution (Spreij et al., 2020). With this information, it can be derived that mode of administration does not impact normative data for TMT B. However, this differs from its counterpoint TMT A because this part does require new normative data.

As mentioned above, Bracken et al. (2018) assessed the TMT adapted for the iPad by Parker-O'Brien to assess reliability and validity. In regards to TMT B, test-retest reliability in three groups produced acceptable values (*r* ranged from 0.33 – 0.80; Bracken et al. 2018). Unlike TMT A, TMT B showed significant difference for mode of administration (TMT B, F (3, 73) 1 = 414.15, $p < .001$, partial n^2 1 = 4 0.37; iPad-TMT B, F (3, 73) 1 = 4 9.44, $p < .001$, partial = n^2 1 = 4 0.28; Bracken et al., 2018). On TMT B, left handers performed slower on the traditional version. This difference in handedness further proves additional norms will be needed for iPad versions of TMT. Although TMT B was able to show adequate test-retest reliability, it was unable to show equivalence when comparing versions. This is consistent with the research mentioned above as it was also unable to prove equivalence between digital versions and traditional versions of that TMT B.

COWAT. Vermeent et al. (2020) administered a verbal fluency task with both sematic and phonemic fluency as part of a larger digital battery to examine if the digital version of the paper pencil tasks load on the same cognitive factors by using the neuropsychological consensus model. However, there were no differences in the administration of these tasks with the iPad. Vermeent et al. (2020) found that both phonemic and semantic fluency tests loaded on the executive functioning factor (z = 6.32, $p > 0.001$ and z = 7.18, $p > 0.001$, respectively).

Tower of Hanoi. Mataix-Cols and Bartres-Fza (2002) analyzed mode of administration for the Tower of Hanoi (ToH) puzzle to assess equivalence. The computerized version used a Ford disc where participants were asked to drag the discs to the different pegs by using their mouse and data was collected automatically; the traditional version data was collected automatically, and participants were required to move the discs to the pegs with their hands. Mataix-Cols and Bartres-Fza (2002) compared 43 undergraduate participants with no history of neurological or psychological disorders on mode of administration. Mataix-Cols and Bartres-Fza (2002) found no significant differences across all variables (total moves, errors, revisions, time). Additionally, the researchers examined learning across mode of administration and found no learning effect from the first to second administration. (Mataix-Cols & Bartres-Fza, 2002).

Another study conducted by Noyes and Garland (2003) found differences between mode of administration in the UK. The computer version utilized a 15-inch monitor and version one of that ToH program authored by Franktiske Folber, where participants were seated in front of the computer and provided the same instructions across both computerized and traditional version (Noyes & Garland, 2003). However, the computerized version provided automated scoring whereas in the traditional version, the examiner scored by hand. Noyes and Garland (2003) compared mode administration for successful completion, number of moves, time taken, and time per move. When comparing the traditional and computerized version of the ToH, Noyes and Garland

(2003) found differences with successful completion, time taken, and time per move. Specifically, the computerized version had higher success rates (computer 92% and traditional 87%). There was a significant difference for time taken t (42) = 5.53, *p* = 0.001 as the computer version was faster (M = 289.83, SD = 161.00) than the traditional versions (M = 476.39, SD = 238.23; Noyes & Garland, 2003). A similar result was seen for time per move, as the computer version was faster with a mean of 5.37 (2.34) compared to the traditional version mean of 10.04 (5.13), resulting in a significant difference (t (42) = -6.85, *p* = 0.001; Noyes & Garland, 2003). Although not significant, a greater number of moves were used to successfully complete the problem on the computerized version (M = 54.43, SD = 22.21) versus the traditional version (M = 49.36, SD = 21.14; Noyes & Garland, 2003). Similar differences were found in a study conducted by Salnaitis et al. (2011) as they found poorer performance on the computer version which was associated with an increase in impulsive responding. However, in another study conducted by Williams and Noyes (2007) where they compared 60 healthy younger adults on the ToH task with both the manual and computer versions, found no significant differences in administration modality. Williams and Noyes (2007) used the same version of the ToH as Noyes and Garland (2003). However, there was a significant finding in regards to amount of time it took as the computer version participants were significantly faster (F (2,54) = 50.45, p < 0.001; Williams & Noyes, 2007). Williams and Noyes (2007) hypothesized that this was to related to working memory as the computer version may reduce working memory load for participants.

Although these two studies are showing inconsistent results in regard to equivalency across computerized versions, it should be noted that they are using different versions of the computerized program; as such, further research needs to be conducted in in clinical populations and with larger sample sizes.

WCST. Unlike Vermeent et al. (2020), Spreij et al. (2020) used the Wisconsin Card Sorting Test (WCST) in their research as part of the d-NPA to assess if normative data is acceptable or equivalent to the tablet version. The iPad version of the WCST had some modifications in comparison to the traditional manual version. Specifically, the cards are presented virtually, and feedback is provided to the patient visually instead of verbally (Spreij et al., 2020). The iPad version has automated scoring (Spreij et al., 2020). Table 3 shows the percentage of participants who had abnormal performances on the variety of scores for the WCST. As such, this table indicates that both number of completed categories and failure to maintain set had more than 10% of the participants in the healthy control group perform below the 10th percentile or had an abnormal performance. This indicates that although for many of the other scores the normative data for paper pencil WCST may be acceptable, for two very important scores, new normative data is indicated.

Table 3

Percentage of Participants Showing in Abnormal Performance

	Stroke n = 56		TBI n = 61		Healthy controls n = 159	
Outcome measures	%	n	%	n	%	n
WCST Total errors	16.4	55	6.8	59	7.6	157
WCST Perseverative errors	9.1	55	6.8	59	4.5	157
WCST Non-perseverative errors	14.5	55	6.8	59	9.6	157
WCST Number of completed categories	16.4	55	16.9	59	12.7*	157
WCST Failure to maintain set	22.2	54	18.6	59	18.6*	156

* Indicates higher than 10% of participants performed below the 10th percentile.

Feldstein et al. (1999) also compared the manual and computer versions of the WCST in 88 student participants and split the participants into four groups: mouse click computer version, mouse auto computer version, keyboard computer version, and touch screen computer version. An additional group of 22 participants was administered the manual version of the WCST. All of these groups were considered equivalent for age education and IQ and were primarily female (Feldstein et al., 1999). Feldstein et al. (1999) used the manual WCST normative data across all groups and compared the following outcome measures: categories completed, total correct, total errors, perseverative errors, non-perseverative errors, and failure to maintain set. The computerized versions were similar to the manual version; however, in the mouse click version, the participant was required to click the next button in order to obtain their next card (Feldstein et al. 1999). The next card was automated in the versions of mouse auto, keyboard, and touch screen (Feldstein et al. 1999). The computerized version provided visual written feedback of “incorrect” and “correct” unlike the manual version that provided verbal feedback (Feldstein et al., 1999). Feldstein et al. (1999) found no differences between the manual version, mouse click version, and mouse auto version. However, they did find significant differences when comparing keyboard version and touch screen version. Specifically, the keyboard version had higher rates of total errors, perseverative errors, non-perseverative errors, and failure to maintain set (Feldstein et al. 1999). The touch screen version had a higher rate of perseverative errors when compared with the manual version (Feldstein et al., 1999). Feldstein et al. (1999) assessed the shape of the distribution using the K-S test for two independent samples. Feldstein et al. (1999) found that the manual version was more negatively skewed than the computer versions (mouse click D = .818, p = .0005; mouse auto D = .864, p = .0005; keyboard D = .591, p = .001; and touch screen, D = .636, p =.0005) for categories completed (Feldstein et al., 1999). Additionally, failure to maintain set resulted in a significant finding; the manual version was more positively skewed than the computerized versions (mouse click D = .682, p = .0005; keyboard D = .591, p = .001; and touch screen, D = .545, p = .0003; Feldstein et al., 1999). Although at first glance there does not appear to be a significant difference between versions, when the scores are standardized to Z scores, the results indicated there is a significant difference for mood administration with the WCST. This result was further examined and confirmed by another study done by Steinmetz et al. (2010) who also compared healthy adults on the manual and computer version on the variance of mean and standard deviation scores on both modes of administration. Steinmetz et al. (2010) found that the percentage of errors in perseverative errors variance was smaller for the computer version; however, failure to maintain sets variance was larger for the computer version. Lastly, in contrast, Wagner and Trentini (2009) found no differences between mode of administration for the WCST in 54 older adults with no neurological difficulties. Specifically, the computer version of the WCST utilized in this study used the keyboard response and compared older adults with the manual version (Wagner & Trentini, 2009). Of note, the study was conducted in Brazil and the groups were considered equivalent for age, education, and MMSE (Wagner & Trentini, 2009).

Table 4

Mean and Standard Deviation for Mode of Administration for the WCST

	Computer		Manual		*p* - Value
	M	SD	M	SD	
WCST Total Number Correct	16.4	55	6.8	59	7.6
WCST Perseverative errors	9.1	55	6.8	59	4.5
WCST Percent concept	14.5	55	6.8	59	9.6
WCST Number of completed categories	16.4	55	16.9	59	12.7*
WCST Failure to maintain set	22.2	54	18.6	59	18.6*

* Indicates higher than 10% of participants performed below the 10th percentile.

Memory

List Learning. Vermeent et al. (2020) administered the Rey Auditory Verbal Learning Test (RAVLT) as part of a larger digital battery to examine if the digital version of paper pencil tasks load on the same cognitive factors using the neuropsychological consensus model. The RAVLT contains scores for learning trails and delay recall and was administered using the iPad. The iPad version provided automated presentation of the list (Vermeent et al., 2020). As to be expected, both the RAVLT learning trails and delay recall loaded on the memory measure ($z = 6.18$, $p < 0.001$ and $z = 6.00$, $p < 0.001$, respectively; Vermeent et al., 2020). Spreij et al. (2020) administered the RAVLT as part of a larger d-NPA to assess if normative data is acceptable or equivalent to the tablet version. The iPad version of the RAVLT was the same version from the research conducted by Vermeent et al. (2020). Table 5 shows the percentage of participants who had abnormal performances on the variety of scores for the RAVLT. As such, this table indicates that immediate recall, delayed recall, and recognition had more than 10% of the participants in the healthy control group who performed below the 10th percentile. This indicates that for the main scores from the RAVLT that new normative data is necessary in order to accurately administer this assessment using a tablet.

Table 5

Percentage of Participants Showing in Abnormal Performance

	Stroke n = 56		TBI n = 61		Healthy controls n = 159	
Outcome measures	%	n	%	n	%	n
RAVLT Immediate recall	44.6	56	41.7	60	33.8*	157
RAVLT Delayed recall	35.7	56	25.0	60	22.9*	157
RAVLT Delayed recall corrected	7.1	56	11.7	60	6.4	157
RAVLT Recognition	12.5	56	16.7	60	11.4*	157

* Indicates higher than 10% of participants performed below the 10^{th} percentile.

Visual Memory. Vermeent et al. (2020) administered the ROCFT immediate recall as part of a larger digital battery to examine if the digital version of paper pencil tasks load on the same cognitive factors using the neuropsychological consensus model. Similar to the ROCFT copy, the immediate recall was administered using an iPad where the participants were asked to draw the complex figure from memory and all drawing was done on the iPad; however, scoring was the same as paper pencil tasks (Vermeent et al., 2020). The ROCFT immediate recall loaded on the memory factor (z = 9.32, $p < 0.001$) which was to be expected (Vermeent et al, 2020). Similar to the research conducted by Vermeent et al. (2020), Spreij et al. (2020) used the ROCFT to assess if current normative data or traditional paper pencil version is comparable or acceptable for the tablet version. Spreij et al. (2020) administered both the immediate and delayed recall trials on the iPad. 12.7% of stroke patients and 18% of TBI participants performed below the 10th percentile and within expectations, while 8.8% of the healthy controls performed below the 10th percentile for immediate recall (Spreij et al., 2020). For delayed recall, 14.5% of stroke participants, 18% of TBI participants, and 9.4% of the healthy control participants performed in the abnormal range or below the 10th percentile. As such, for both immediate and delayed recall for the ROCFT, normative data appears to be acceptable for the tablet version. However, as stated above, the ROCFT copy normative data may not be applicable for the tablet version; thus, for clinical use, it may be helpful to derive new normative data.

Eligibility Criteria

What Does the Neuropsychologist Need?

Similar to paper pencil or face to face neuropsychological testing, testing conducted through the use of a computer or iPad comes with an additional set of considerations. Many of those considerations surround the technology needed in order to properly conduct computer neuropsychological assessments that will be discussed in addition to considerations that need to be taken in regard to best fit of population or comfort level with computer or iPad use.

With computerized assessments, consideration should be taken into whether it is Internet based or is a software download. Primarily, the studies reviewed utilized a software downloaded onto either an iPad or computer. Vermeent et al. (2020) and Spreij et al. (2020) utilized a Tablet – iPad with an Apple pencil that was set to screen size of 12.0 resolution of 2731 x 2048 pixels; both researchers reported using an Apple iPad Pro. Additionally, both Vermeent et al. (2020) and Spreij et al. (2020) used d-NPA research prototype by Phillips research that provided all-digital versions of the assessment. Bracken et al. (2010) used an iPad and TMTs were administered on Apple iPad-Air (9.7-

inch diagonal LED-backlit Multi-Touch display with IPS technology) with a resolution of 2048 x 1536 and iOS version 8.4 utilizing the auto-brightness setting. Cernich et al. (2007) discussed the operating system necessary for running software for neuropsychological assessments, specifically, operating systems typically have 15ms to 55ms delay in display rates which can impact the reliability of the measure and add a new source of error in testing. Cernich et al. (2007) discussed the programs and operating systems that can impact timing because different operating systems' time is based off a software clock or a system clock. The current gold standard is to use a real time operating system that requires specialized hardware that can be expensive because it increases the accuracy of timing (Cernich et al., 2007).

To properly use software and web based computerized assessment, it is the clinician's responsibility to ensure they obtain detailed technical information from publishers including hardware/software specifications and timing resolution information because this can impact the assessments' ability to run correctly (Cernich et al., 2007). Therefore, when a clinician is choosing to use a computerized assessment, it is of great importance for the clinician to make sure that their current operating system, hardware, and software match procedures provided in the computerized assessment manuals.

What Populations are Best Suited for Computerized Assessments?

Demographics. In the many previously mentioned studies, the vast majority utilized healthy controls of young adults. Specifically, the typical age range for the studies was from 18 to 60 (Aşkar et al., 2012; Bracken et al., 2018; Edwards et al., 1996; Feldstein et al., 1999; Noyes & Garland, 2003; Salanaitis et al., 2011; Steimnetz et al., 2010; Williams & Noyes, 2007). However, a few studies included older adults. Dion et al. (2020), Vermeent et al. (2020), Spreij et al. (2020), and Saxton et al. (2009) only used adults aged 65 and above. All the studies had a relatively equal number of men and women. Similar to telephone neuropsychological assessments, little consideration was given to ethnic or racial monitories because the majority of the participants included were White.

Cognition. Similar to demographic information, the vast majority of the studies utilized healthy controls in their comparison studies (Aşkar et al., 2012; Bracken et al., 2018; Edwards et al., 1996; Feldstein et al., 1999; Noyes & Garland, 2003; Salanaitis et al., 2011; Steimnetz et al., 2010; Williams & Noyes, 2007). However, a select few studies did use older adults with the intention to compare cognitive status. Dion et al. (2020), Spreij et al. (2020), Saxton et al. (2009), and Vermeent et al. (2020), utilized older adults with varying cognitive difficulties. Specifically, Spreij et al. (2020) included older adults with traumatic brain injuries, stroke, and healthy older adults. Saxton et al. (2009) only used older adults with either MCI or normal cognition. Dion et al. (2020) and Vermeent et al. (2020) only utilized older adults with normal cognition.

Exclusion Criteria. Although the psychologists may want to assess a participant's ability to use a tablet or iPad, Spreij et al. (2020) found no significant differences between the effect of tablet familiarity on test performance because this was done by running a hierarchical method with predictors for age, sex, and level of education and a second model was run for iPad familiarity to assess for improvement of Model 1 two model 2 by looking at the F-change. However, other articles looked at experience with computers in healthy young adults. Williams and Noyes (2007) found significant differences between experience in novice computer users specifically those with experience had higher scores and faster responses on the Tower of Hanoi task. Both studies conducted by Spreij et al. (2020) and Vermeent et al. (2020) included the following exclusion criteria: healthy older adults with no hearing or vision issues unless corrected, those with severe communication, motor, neurological, or psychiatric disorders. Additionally, participants were not included if they were unable to use a tablet or perform digital tests. Dion et al. (2020) and Saxton et al. (2009) also excluded those whose first language was not English.

Implications for Clinical Practice

Benefits of Computerized Assessments

As technology advances, so does the ability to use this technology as an advantage when conducting neuropsychological assessments. A review of computerized assessment conducted by Zygouris and Tsolaki (2015) listed the following benefits of computerized assessments: efficiency, increased reliability with scoring, additional scores for reaction time and impulsivity, accurate recording of responses, and the ability to automatically store and compare a person's performance over time. Another review conducted by Noyes and Garland (2008) indicated many of the same benefits but also reported on the increased standardization of test environment in test instructions. Specifically with computerized assessments, they are able to present the information in the same way and at the same speed each time thus decreasing errors in administration (Noyes and Garland, 2008). Additionally, Saxton et al. (2009) found that the CAMCI had better sensitivity and specificity with MCI. Dion et al. (2020) found with that the dCDT score TCT was able to show subtle changing in MCI. Lastly. Spreij et al. (2020) provided practicians with questionaries to better understand their experience. Spreij et al. (2020)

found that 91% of the participants reported performing the tests on a tablet to be pleasant. In regard to visibility of tests, participants drawing and the ability to draw on the tablet were considered satisfactory by both stroke and TBI in healthy control participants (Spreij et al., 2020). In regards to drawing, generally participants with stroke or TBI had more positive responses than the healthy controls (Spreij et al., 2020).

Disadvantages of Computerized Assessments

As with all advantages, there are always disadvantages. Although computerized neuropsychological assessments can streamline the assessment process, there also can be many negative implications to using computerized assessment. Specifically, if when much care and attention is given to the computer hardware and software, this does not always mean it is going to work perfectly. Computers are fallible and as such can freeze or crash during testing, which can impact the time allotted to finish the assessment and the participants' ability to accurately complete the assessment (Noyes & Garland, 2008). Other disadvantages include increased eyestrain due to the computer screen, possible concerns with confidentiality if using web-based assessments, and increased difficulty with those who have minimal computer skills (Noyes & Garland, 2008). Zygouris and Tsolaki (2015) discussed possible impacts of computerized assessment with participants who lack knowledge or have limited experience with computers. This was further substantiated by Williams and Noyes (2007) in that novice computer users performed worse and had increased performance time on the ToH task.

Spreij et al. (2020) assessed the feasibility of the d-NPA and found that 94% of the participants were able to complete the entire assessment. However, one stroke patient had difficulties with the ROCFT and was unable to complete the Stroop or WCST; the participant reported to be too tired (Spreij et al., 2020). Additionally, one TBI patient was unable to complete four of the tests as they reported sensory overload (Spreij et al., 2020). Furthermore, three patients had difficulty with the brightness and needed a reduction of brightness, volume adjustment, or an extra break (Spreij et al., 2020). Specifically, 5% of the participants needed an extra break and 6% needed adjustments on the iPad (Spreij et al., 2020). Four participants reported their experience of tablet administration to be very unpleasant; they reported sensory overload and felt the administration mode was more tiring and required more mental energy (Spreij et al., 2020). Furthermore, precipitants had difficulty with the surface of the tablet as they felt it gave less friction, they felt the tablet screen was less accurate, and frustration as errors could not be erased on the tablet (Spreij et al., 2020). Others experienced difficulties with inability to rest their hand on the tablet and they felt their hand was in a different position when using the Apple pencil (Spreij et al., 2020).

Missing Pieces

Many of these research articles do an excellent job of outlining the comparisons between modes of administration; however, little is done to provide guidance to a practicing neuropsychologist. Specifically, Spreij et al. (2020) used healthy controls and effectively found that 34% had an abnormal performance, which questions the acceptability of paper pencil normative data for computer use. Furthermore, many of the studies for the ToH, Stroop, and WSCT found significant differences in means between modes of administration that again questions the use of current normative data for computerized assessments. This questions further the test’s ability to indicate impaired performance because computer testing may have more false positives of impairment.

Additionally, it was noted frequently that those who participated in computerized assessments often engaged in increased impulsive behavior leading to increased error rates and faster responding.

Lastly, the above the articles do not address clinical implications such as report writing and behavioral observations. Oftentimes, many of the assessments required modifications to the original test that typically would be described in the report as they may impact that as a whole. If the clinician is providing the computerized assessment where the participant or patient is alone, this decreases a clinician's ability to obtain robust behavioral observations that can impact one’s understanding of the person's performance.

Emerging Technology

Virtual Reality Assessments

Parsey and Schmitter-Edgecombe (2013) conducted a review on both computerized assessments and virtual reality neuropsychological assessments. Virtual reality assessments were initially developed for integrating computerized versions of traditional paper pencil tests into a virtual environment to obtain both behavioral and cognitive information that would go beyond evidence typically obtained in traditional paper pencil tasks (Parsey & Schmitter-Edgecombe, 2013). As such, this would allow a neuropsychologist to be able to observe a patient's approach to daily tasks in stimulated environments that would better represent everyday life and increase ecological validity in neuropsychological assessments (Parsey & Schmitter-Edgecombe, 2013).

Virtual reality tasks can be used to assess cognitive domains including attention, memory, and executive functioning; however, when a task is implemented into virtual reality, this can fundamentally change the intention of the task and as such can impact the cognitive construct meant to be measured (Parsey & Schmitter-Edgecombe, 2013). Many virtual reality tasks are expanding cognitive information being able to be assessed including multitasking components and higher order tasks that can fresh tap into multiple cognitive domains at once (Parsey & Schmitter-Edgecombe, 2013). This is frequently done by having participants engage in activities of daily living through a virtual reality environment. Most frequently, stroke populations, Parkinson's disease, and TBI patients are asked to engage in assessments including “look for a match," a virtual reality Stroop task, virtual reality paced serial assessment test, virtual reality cognitive performance assessment test, virtual classroom, virtual errands test, virtual multiple errands, and virtual kitchen. Parsey and Schmitter-Edgecombe (2013) found that virtual reality cognitive assessments can help identify cognitive deficits. Driving simulation test have been utilized to better understand cognitive demands used during driving with many clinical populations. Currently, research is divided as some studies have provided support for virtual reality driving simulations while other driving simulators may not be completely adequate (Parsey & Schmitter-Edgecombe, 2013). Much of the research in the review conducted by Parsey and Schmitter-Edgecombe (2013) found a strong base for utilizing virtual reality assessments; however, the research indicates this should be used in conjunction with traditional neuropsychological assessments especially when competency decisions are in question.

Clinical Pearls

Table 6 provides a summary of the previously mentioned research regarding validity and modifications, and if new normative data is need for each assessment reviewed. Table 6 can be used as a quick reference guide for clinicians when deciding what assessments to utilize during remote computerized assessments.

Table 6

Does Mode of Administration Matter for Computerized Assessments

Test	Valid	Modifications	New normative data needed
CAMCI	Yes	Yes	New normative data was developed for this test
dCDT	Yes	Yes	New normative data was developed for this test
Digit Span	Yes	No	Lezak's distribution found the tests to be equivalent; however, no mean or standard deviations were assessed
TMT A	Variable	No	Yes, Lezak's distribution found the tests to be not equivalent and difference in handiness was noted
Stroop	Yes	No	Differences in mean scores found as participants were faster on the computer version
Cancellation	Yes	No	Lezak's distribution found the tests to be equivalent; however, no mean or standard deviations were assessed
Rey-O	Yes	No	Lezak's distribution found the tests to be equivalent; however, no mean or standard deviations were assessed
Cube Drawing	N/A	No	Yes, Lezak's distribution found the tests to be not equivalent and no mean or standard deviations were assessed
Line Orientation	N/A	No	Differences in mean scores found as participants had higher scores on paper pencil version
TMT B	Yes	No	Differences in mean scores and difference in handiness was noted
COWAT	Yes	No	No analyses were conducted for mean or standard deviation difference
Towe of Hanoi	Yes	No	Depends on the program being used but mean differences were found
WCST	Yes	No	Yes, differences were found between computer and tradition versions
List learning	Yes	No	Yes, Lezak's distribution found the tests to be not equivalent and no mean or standard deviations were assessed
Visual Memory	Yes	No	Lezak's distribution found the tests to be equivalent; however, no mean or standard deviations were assessed

*N/A= not applicable

CHAPTER IV: TELEHEALTH: TELEVIDEO NEUROPSYCHOLOGICAL ASSESSMENTS

Neuropsychological assessments can be administered in various ways and, as such, there have been many research studies conducted to compare modes of administration. This chapter will review three of the main research studies that compare face to face traditional neuropsychological assessments with video teleconference assessments. Unlike the telephone neuropsychological administration research, many of these studies have conducted research to compare mean differences and center deviation differences between mode of administration. Additionally, a recent critical review was conducted to assess the validity of televideo neuropsychology assessment in response to COVID-19. Additionally, this chapter will provide neuropsychologists pertinent information on populations best suited for televideo assessments, technology needed, and considerations for both diagnosis and report writing.

Overview of Current Research

Cognitive Screeners

MMSE. Munro Cullum et al. (2014) examined the validity of video teleconference neuropsychological assessment through the use of a brief battery and compared face to face (FTF) assessment with a video teleconference (VTC) assessment. Two hundred and two older adult participants with either MCI, probable AD, and or normal cognition were split into two groups, VTC or FTF. Munro Cullum et al. (2014) administered the Mini- Mental State Examination (MMSE), Hopkins Verbal Learning Test-Revised (HVLT-R), Digit Span forward and backward, short form Boston Naming Test (BNT), Letter and Category Fluency, and Clock Drawing. Test administration was conducted using telehealth clinics in both rural and urban areas where local staff seated participants at the computer screen but were not present during VTC. The following analyses were conducted to determine the validity of VTC neuropsychological test administration: Intraclass Correlations Coefficients (ICC), the Bradley-Blackwood Procedure to examine mode of administration bias and if there was a significant result a paired *t*-test was done for means and Pitman test for variances, and Bland-Altman plots. Munro Cullum et al. (2014) found no differences on all analysis between FTF (M = 27.6, SD = 3.09) and VTC (M = 27.6, SD = 3.10) for the MMSE as p values were greater than 0.05. Additionally, the ICC was 0.905 with a ($p < 0.0001$; Munro Cullum et al., 2014). This indicates that the MMSE FTF administration is highly comparable to VTC administration as no mean or standard deviation differences were found and they were considered to be correlated with each other.

Another study conducted by Montani et al. (1996) also administered the MMSE to six women and four men were administered MMSE via videophones. Montani et al. (1996) compared participants' performance on mode of administration for both the MMSE and Clock Draw. Montani et al. (1996) reported for the video phone condition that each room had a camera, television screen, and microphone with the clinician who operated the mobile camera and another clinician who was in the other room. Both computers were connected via a coaxial cable and they reported no changes to test administration. Unlike previous studies mentioned above, this study found decreased performance when comparing the use of videophones versus FTF with a ($p = 0.008$; Montani et al., 1996). Specifically, the researchers did mention difficulty with hearing and decreased communication in the video phone condition that may have impacted participants' ability to hear and pay attention properly.

A systematic critical review was conducted by Marra et al. (2020) on 19 articles compared FTF and VTC neuropsychological assessments to determine validity in equivalence between mode of administration. The critical review found the MMSE to be a valid videophone psychological assessment for older adults in that there were nine out of 10 articles that found no difference in mean scores for mode of administration. However, no information was provided regarding Bland-Altman plots and standard deviation differences.

MoCA. Chapman et al. (2019) administered the MoCA to 48 stroke survivors from Australia using a crossover design. All participants were administered the MoCA in both the FTF and VTC neuropsychological assessment conditions approximately two weeks apart. To compare differences in mode of administration, a repeated measures *t*-test, ICC, Bland-Altman plot and multivariate regression modelling were used (Chapman et al., 2019). Chapmen et al. (2019) used a cloud-based video conferencing site to administer the MoCA remotely and provided only the visuospatial, executive functioning, and naming items in person that were in an envelope. Additionally, no changes in administration were made. Chapman et al. (2019) found no significant differences in scores across mode of administration $t(47) = .44$, $p = .658$. However, there were wide limits of agreement in the Bland-Altman plots that can indicate inconsistency in mode of administration and a weak ICC as the (ICC = 0.615; Chapman et al., 2019). Given this information, mode of administration does impact psychometric similarity between FTF and VTC with the MoCA.

Hewitt and Loring (2020) also administered the MoCA with modifications; however, this article did not provide any statistical analysis in regard to comparisons or equivalents for mode of administration. Hewitt and Loring (2020) presented visual stimuli individually to enhance attention and utilized notice the screen grab feature to obtain the participant's cube copy and clock draw. Additionally, they requested the patient close their eyes to limit distractions and cheating for orientation questions (Hewitt & Loring, 2020).

The systematic critical review conducted by Marra et al. (2020) found four studies that compared mode of administration on the MoCA. These studies indicated strong reliability metrics as the ICC ranged from .59 to .93 across the four studies and no study found mean differences between FTF and VTC. Again, there was no mention of Bland-Altman plots or standard deviation differences. However, Chapman et al.'s (2019) study found wide limits of agreement in the Bland-Altman plot indicating that mode of administration does matter.

RBANS. Galusha-Glasscock et al. (2016) conducted a study to examine mode of administration effects for the Reparable Battery for the Assessment of Neuropsychological Status (RBANS) in 18 older adults. This study sample included seven cognitively normal adults, six adults diagnosed with MCI, and five adults with a diagnosis of AD (Galusha-Glasscock et al., 2016). Specifically, each participant was administered the RBANS in both the FTF and the VTC mode of administration using alternate forms of the assessment (Galusha-Glasscock et al., 2016). Of note, a Polycom iPower 680 series video conferencing system was used that allowed the examiner to adjust the camera to see both the participant and the stimuli simultaneously (Galusha-

Glasscock et al., 2016). Galusha-Glasscock et al. (2016) provided an assistant at the beginning of testing to provide each participant with an introduction regarding assessment procedures, explanation of materials, and an explanation of the TV monitor for the VTC condition. Additionally, the accommodations used for the VTC condition included the following: the examiner held up the stimulus book for figure copies, line orientation, picture naming, and coding and a blank piece of paper and pen were available in the room for participants as well as a copy of the coding form from the test protocol (Galusha-Glasscock et al., 2016). Galusha-Glasscock et al. (2016) conducted an interclass correlations and a paired-sample *t*-test to compare RBANS index scores for motive administration conditions across the whole sample. Table 7 provides both descriptive statistics and the interclass correlation results for all indices scores for the RBANS. As Table 7 shows, there is moderate to strong significant interclass correlations for each index. However, Galusha-Glasscock et al. (2016) did not provide *p* values for the paired-samples *t*-test but did report no significant differences for the means comparing mode of administration for both FTF and VTC. Galusha-Glasscock et al. (2016) indicated similar means, but no information was provided regarding standard deviation differences and or Bland-Altman analysis.

Table 7

Descriptive Statistics and Correlations for Mode of Administration for the RBANS

Index	Face to face M(SD)	Video Teleconference M(SD)	ICC (r)
Immediate memory	97.17(27.12)	96.22 (24.06)	.84**
Visuospatial/constructional	94.89 (20.16)	92.72 (23.00)	.59*
Language	95.94 (13.49)	95.56 (10.60)	.75**
Attention	96.33 (18.69)	93.33 (16.80)	.81**
Delayed memory	90.83 (30.37)	93.28 (27.06)	.90**
Total scale	94.50 (23.10)	93.06 (19.74)	.88**

*Significant at the $p < 0.01$
**Significant at the $p < 0.001$

Attention and Working Memory

Digit Span. As stated previously, Munro Cullum et al. (2014) examined the validity of video teleconference neuropsychological assessment through the use of a brief battery and compared face to face (FTF) assessment with a video teleconference (VTC) assessment. Munro Cullum et al (2014) administered a Digit Span test to 202 older adult participants with either MCI, probable AD, and or normal cognition that were split into two groups, VTC or FTF. Munro Cullum et al. (2014) did not note any changes to administration for the Digit Span test. Statistical analysis did not note any differences between FTF administration and VCT administration as the p value was greater than 0.05 for both digits span forward and digit span backwards. The interclass correlation was moderate for both digit span forward and backwards (0.590 and 0.545 respectively; Munro Cullum et al., 2014).

Grosch et al. (2015) conducted a study to determine equivalence for mode of administration on three brief neuropsychological assessments. Grosch et al. (2015)

administered the digit span test according to standard administration procedures to eight older adults. Grosch et al. (2015) utilized the Bradley-Blackwood Procedure and the ICC to assess equivalence between mode of administration. There was no significant difference in means or standard deviations between (FTF [M = 9.88, SD = 2.17] and VTC [M = 9.63, SD = 2.77] as p = 0.946; Grosch et al., 2015). Additionally, the ICC was considered strong, (ICC = 0.72 Grosch et al., 2015).

Jacobsen et al. (2003) conducted a study with 32 healthy controls who were split into either the FTF or VTC condition to assess equivalence and reliability for mode of administration on neuropsychological assessments. There was no report of modifications for the administration of digit span and standard procedure was utilized. Reliability coefficients were utilized to examine consistency of scores across mode of administration and paired *t*-tests were used to assess mean differences. There was no significant difference in means when comparing (FTF [M = 11.8, SD = 1.8] and VTC [M = 12.1, SD = 2.2] with p = .33; Jacobsen et al., 2003). FTF and VTC were found to be strongly correlated with each other (r = .82; Jacobsen et al., 2003).

The systematic critical review conducted by Marra et al. (2020) found six studies that compared mode of administration on the Digit Span Test. These studies indicated fair validity metrics as the ICC was generally in .50 range. Marra et al. (2020) reported on a study that did find a significant difference between means that was the study conducted by Wadsworth et al. (2018). Specifically, Wadsworth et al. (2018) reported mean differences between FTF 5.9 (1.4) and VTC 5.5 (1.3) with at p = 0.004 for only Digit Span Forward but did not report any difference for Digit Span Backwards. Given this review article, there appears to be differences in validity statistics based on if articles are reporting Digit Span Total versus information for digits Span Forwards and Backwards because there is better reliability metrics when digit span total is provided. Additionally, no information regarding standard deviation differences or Bland-Altman plots are discussed in this critical review.

Seashore Rhythm Test. As stated above, Jacobsen et al. (2003) conducted a study to assess reliability and equivalence for mode of administration on neuropsychological assessments. The Seashore Rhythm Test was administered as part of a larger neuropsychological battery. Jacobsen et al. (2003) did not report any modifications in test administration. There was a significant difference in means when comparing FTF (M = 26.8, SD = 2.8) and VTC (M = 27.6, SD = 2.1) as (t = 2.37, p = .0.3; Jacobsen et al., 2003). FTF and VTC were found to be strongly correlated with each other (r = .77; Jacobsen et al., 2003). Upon exit interviews, researchers were informed by participants that they felt they could focus better during VTC due to perceived reduced distractions (Jacobsen et al., 2003). Given this information, the seashore rhythm test needs further evaluation to determine equivalence between mode of administration.

Graphomotor Speed

As previously stated, Jacobsen et al. (2003) conducted a study to assess reliability and equivalence for mode of administration on neuropsychological assessments. Jacobsen et al. (2003) administered the Groove Pegboard test via FTF instructions. However, during the remote session, the test was demonstrated with a document camera and the participant was provided with their own Groove Pegboard. Instructions were initially demonstrated with the document camera and then the participant used the corresponding objects that were located on the desk in front of the participant. There was not a significant difference in means when comparing FTF and VTC for both nondominant and dominant hands as the P value was greater than 0.2 (Jacobsen et al., 2003). FTF and VTC were found to be strongly correlated with each other as $r > .7$ for both the nondominant in dominant hand conditions (Jacobsen et al., 2003). Although these researchers were able to determine equivalent through means and correlation, there still is question on how Groove Pegboard would be administered in a typical clinical setting via VTC. Therefore, further research needs to be conducted into the feasibility administering the Groove Pegboard test remotely.

Processing Speed

Symbol Digit Modalities Test. As mentioned, Jacobsen et al. (2003) conducted a study to assess reliability and equivalence for mode of administration on neuropsychological assessments. Jacobsen et al. (2003) administered the Symbol Digit Modalities Test (SDMT) with no modifications in instructions and was initially demonstrated with the document camera and then the participant used the corresponding objects located on the desk in front of the participant. There was not a significant difference in means when comparing FTF and VTC for both the oral and written versions, the P value was greater than 0.8 (Jacobsen et al., 2003). FTF and VTC were found to be moderately correlated with each other as $r = .69$ for the written (Jacobsen et al., 2003). However, the oral version had a weak correlation as ($r = .37$; Jacobsen et al., 2003). Jacobsen et al. (2003) hypothesize this was due to the short amount of time between tests. Although Jacobsen et al. (2003) were generally able to establish equivalence between means, there was no discussion regarding standard deviation variances or Bland-Altman plots that would further strengthen psychometric similarity between mode of administration.

Trail Making Test. Wadsworth et al. (2016) conducted a study to determine feasibility and equivalent of mode of administration with 84 participants who had a diagnosis of MCI, dementia or cognitively normal and were from the Choctaw Nation. Wadsworth et al. (2016) did not report any procedural changes in the instructions provided in the VTC condition. To determine feasibility and reliability, both interclass correlations (ICC) and paired samples T-test were used. For Oral Trails A, there was a significant difference for means between FTF 8.9 (2.4) and VTC 11.1 (3.0) as ($t = -9.60$, $p < 0.001$; Wadsworth et al., 2016). However, there was no significant difference in means for Oral Trails B between FTF 76.0 (90.5) and VTC 78.8 (77.2) as ($t = -0.35$, $p = 0.726$; Wadsworth et al., 2016). The ICC was found to be significant for both Oral Trails A and B with an (ICC of 0.83 and 0.79; Wadsworth et al., 2016). This article is an excellent first step at determining feasibility and reliability for oral trail making test; however, further investigation needs to be completed as no Bland Altman plots were use and no analysis was conducted regarding standard deviations, especially given the large difference between Oral Trails B standard deviations.

Although no statistical analyses were run by Hewitt and Loring (2020), they did report using the oral version of Trail Making Test that derived from the adaptation from Mrazik et al. (2010). It should be noted that they did make some administration changes; these researchers provided the visual stimuli on the computer screen via zoom screen share as it was scanned into the computer (Hewitt & Loring, 2020). The patient was then requested to respond aloud (Hewitt & Loring, 2020). Additional administration notes included if a patient made an error during the sample trial, the Emory clinic would provide slides that demonstrated the proper order to ensure the individual best understands the task with visual aids (Hewitt & Loring, 2020). If the patient were to make an error during the task, the psychometrist provides a prompt and shows the patient where to start via the cursor on the screen (Hewitt & Loring, 2020). Given the changes in administration for VTC, follow-up research will be necessary to indicate no differences in means and standard deviations along with interclass correlations to provide reliability to substantiate FTF normative data for this mode of administration.

Language

BNT. As mentioned previously, Munro Cullum et al. (2014) conducted a study to determine reliability and validity of the BNT-15 when administered through VTC. Two hundred and two older adult participants with either MCI, probable AD, and or normal cognition were split into two conditions, FTF and the VTC, and were administered the BNT-15. Munro Cullum et al. (2014) did not report any administration changes when administering the BNT-15 in the VTC condition. BNT-15 revealed significant differences on both the Bradley-Blackwood procedure (p = 0.003) and the Pitman test (p = 0.004) when comparing means for FTF (M = 13.3, SD = 2.16) assessment condition and the VCT (M = 13.1, SD = 2.43; Munro Cullum et al., 2014). This indicated that the variances were statistically different when comparing FTF with VCT. However, the Bland-Altman plots showed very low bias; this indicates that the mode of administration is psychometrically similar (Munro Cullum et al., 2014). The interclass correlation analysis was considered strong with an (ICC = 0.812; Munro Cullum et al., 2014). Unlike other assessments in the brief battery used by Munro Cullum et al. (2014), there is some question to this psychometric similarity of BNT-15 when mode of administration changes from FTF to VCT.

Wadsworth et al. (2016) conducted a study to determine feasibility and equivalent of mode of administration with 84 participants who had a diagnosis of MCI, dementia or cognitively normal and were from the Choctaw Nation. No modifications were noted for the BNT and no information was provided regarding procedural instructions for the VTC condition (Wadsworth et al., 2016). As stated above, both ICC and paired samples T-test were used to determine equivalence. There was a significant difference for means between FTF 12.9 (2.2) and VTC 12.5 (2.6) as (t = 3.21, p = 0.002; Wadsworth et al., 2016). The ICC was found to be significant and strong as the (ICC = 0.093; Wadsworth et al., 2016). This article is an excellent first step at determining feasibility and reliability for BNT but further investigation needs to be completed with Bland-Altman plots and determining differences for standard deviations.

The systematic critical review conducted by Marra et al. (2020) found four studies that compared mode of administration on BNT. Only one of these studies found a significant difference between means, which is the study noted above in this section. Additionally, Marra et al. (2020) reported strong reliability statistics for all four studies as the ICC ranged from 0.812 to 0.930. Additionally, no information regarding standard deviation differences or Bland-Altman plots are discussed in this critical review.

WAIS Vocabulary. As mentioned, Jacobsen et al. (2003) conducted a study to assess reliability and equivalence for mode of administration on neuropsychological assessments and administered Vocabulary with no modifications. There was no significant difference in means when comparing FTF (M = 29.6, SD = 4.5) and VTC (M

= 29.5, SD = 4.0) with (*p* = .83; Jacobsen et al., 2003). FTF and VTC were found to be strongly correlated with each other as (*r* = .86; Jacobsen et al., 2003). Jacobsen et al. (2003) was able to establish equivalence between means. There was no discussion regarding standard deviation variances or Bland-Altman plots that would further strengthen psychometric similarity between mode of administration.

Visuospatial Ability

VOSP Silhouettes. As stated, Jacobsen et al. (2003) conducted a study with the intent to assess reliability and equivalence for mode of administration on neuropsychological assessment. Jacobsen et al. (2003) administered Visual Object and Space Perception (VOSP) Silhouette subtest with no instruction modifications and was demonstrated with the document camera then the participant used the corresponding objects located on the desk. No significant difference in means were noted when comparing FTF (M = 11.8, SD = 2.0) and VTC (M = 11.8, SD 2.2) with (*p* = .84; Jacobsen et al., 2003). FTF and VTC were found to be moderately correlated with each other as (*r* = .64; Jacobsen et al., 2003). Although Jacobsen et al. (2003) was able to establish equivalence between means, there was no discussion regarding standard deviation variances or Bland-Altman plots that would further strengthen psychometric similarity between mode of administration.

Rey-O. Hewitt and Loring (2020) published an article regarding the procedures utilized to administer a brief neuropsychological assessment battery through VTC; no statistical analyses were run to assess equivalency. Hewitt and Loring (2020) administered the Rey-O Complex Figure for both copy and memory trials with no modifications in instructions. For the copy trial, the patient was instructed to their folder with paper prior to the start of testing then the patient was asked to fold the paper in half and copy the figure on the screen (Hewitt & Loring, 2020). Once the patient was finished, they were asked to hold the paper up to the screen where the psychometrist would screen grab the copy without capturing the patient's face (Hewitt & Loring, 2020). After the copy was screen grabbed, the patient was asked to place the copy in a folder and asked not to look at it again (Hewitt & Loring, 2020). Hewitt and Loring (2020) utilized a similar procedure for the delayed trail but instead instructed to grab a blank piece of paper and fold it in half to reproduce the original figure they had copied. Again, Hewitt and Loring (2020) utilized the screen grab feature to capture the reproduction for the memory delay.

Executive Functioning

COWAT. As stated previously, Munro Cullum et al. (2014) conducted a research study to examine the validity of VTC administered neuropsychological assessments by administering a short battery in comparing modes of administration. Munro Cullum et al. (2014) administered a FAS and categories fluency to 202 older adult participants with either MCI, probable AD, and or normal cognition that were split into two groups VTC or FTF in order to compare mode of administration. In regard to FAS, there was no difference in mode of administration across all statistical analysis as means for FTF 38.5 (13.48) and VTC 38.0 (13.61) were not statistically different with p values across all statistical analysis or greater than 0.05 (Munro Cullum et al., 2014). Category fluency yielded a similar result as there was no statistical difference between FTF (M = 17.0, SD = 5.50) and the VTC (M = 16.7, SD = 6.06) administration with a p value that was greater than 0.05 (Munro Cullum et al., 2014). Additionally, for both FAS and category fluency, their interclass correlations were considered to be strong as both were greater than 0.7 (Munro Cullum et al., 2014). Similar to previous results, Munro Cullum et al. (2014) was able to establish reliability and validity for both FAS in category fluency for VTC mode of administration. As such, this information provides the neuropsychologist to be able to use current normative data for VTC mode of administration.

Wadsworth et al. (2018) administered a brief neuropsychological assessment battery to 197 older adults who were defined as either impaired or unimpaired to determine psychometric similarity between FTF and VTC. Wadsworth et al. (2018) did not report any modifications with standardized instructions and procedures for the VTC condition. Wadsworth et al. (2018) utilized a repeated measures ANCOVA to compare mode of administration. In regard to FAS fluency, no significant difference was found when comparing means between FTF and VTC (Wadsworth et al., 2018). However, Animal Category revealed a significant difference between means when comparing FTF 18.46 (4.76) with VTC 18.76 (5.07) with a ($p < 0.001$; Wadsworth et al., 2018) although a small effect size was noted with a (Cohen's $d = 0.063$; Wadsworth et al., 2018). A previous study conducted by Wadsworth et al. (2016) to determine equivalence between mode of administration in the Native American population found similar results by Wadsworth et al. (2018). Unlike the newer study, Wadsworth et al. (2016) found no significant differences between modes of administration for both FAS and Animal Categories.

The systematic critical review conducted by Marra et al. (2020) found seven studies that compared mode of administration on FAS and five studies that compared mode of administration on Category fluency. The FAS fluency studies had no mean difference and generally strong reliability statistics as the ICC ranged from 0.83 to 0.93 (Marra et al., 2020). However, the statistics were a bit more variable for category fluency. As noted above, there was one study conducted by Wadsworth et al. (2018) that found mean differences and the reliability statistics were more variable and moderate with an (ICC range of 0.58 - 0.74; Marra et al., 2020).

Clock Draw. Munro Cullum et al. (2014) administered the Clock Draw test to compare mode of administration within their research population. Munro Cullum et al. (2014) had participants hold up their Clock draw after it was complete for examiners to score in real time and after the assessment was complete, all materials were included in a package that was sent back to the examiner. The statistical analysis revealed no significant difference for mode of administration means for FTF 5.6 (0.80) and VTC 5.6 (0.89) with a P value greater than 0.05 (Munro Cullum et al., 2014). The interclass correlations result was considered to be a moderate correlation as the (ICC = 0.709; Munro Cullum et al., 2014). Cullum et al.'s (2014) research indicated there is psychometric similarity between FTF and VCT for the Clock Draw test.

As previously stated, Montani et al. (1996) conducted a study to compare mode of administration with the Clock Draw test. Six women and four men were administered the Clock Draw test via videophones and FTF (Montani et al., 1996). Montani et al. (1996) did not provide information regarding procedures utilized. Unlike previous studies mentioned, this study found a significant difference between means when comparing the use of videophones (M = 19.8) versus FTF (M = 22.4) with a (p = 0.006; Montani et al., 1996). Specifically, the researchers did mention difficulty with hearing that may have led to decreased communication in the videophone condition thereby impacting the participant's ability to hear and pay attention properly.

The systematic critical review conducted by Marra et al. (2020) found eight studies that compared mode of administration on Clock Draw to determine equivalence. Of the eight studies, two reported significant differences between means when comparing FTF with VTC one of which was reviewed previously (Marra et al., 2020). Additionally, there was variable validity statistics (ICC ranged 0.42 – 0.71; Marra et al., 2020) Furthermore, no information regarding standard deviation differences or Bland-Altman plots are discussed in this critical review.

Memory

HVLT-R. As previously mentioned, Munro Cullum et al. (2014) conducted a study with the purpose to examine the validity and reliability of VTC administered neuropsychological assessments. Specifically, the researchers administered the HVLT-R with no modifications noted for VTC (Munro Cullum et al., 2014). The statistical analysis revealed a moderate interclass correlations as the (ICC = 0.709; Munro Cullum et al., 2014). There was a significant difference for mode of administration means for FTF 22.6 (6.98) and VTC 23.4 (6.90) with a $p = 0.005$ for the paired sample *t*-test and a $p = 0.019$ for the Bradley-Blackwood procedure (Munro Cullum et al., 2014). However, the Pitman Test was not significant, and the Bland-Altman plots showed very low to no bias (Munro Cullum et al., 2014). Munro Cullum et al. (2014) reports that this indicates that the mode of administration does not impact the psychometric properties of the test. However, given that there is a significant difference between the means, this may impact current provided normative data.

The systematic critical review conducted by Marra et al. (2020) found five studies that compared mode of administration on HVLT-R to determine equivalence; three studies included HVLT-R delayed recall. Of the five studies, one reported significant difference between means when comparing FTF with VTC that was reviewed previously (Marra et al., 2020). However, the effect size was small in this article (g = 0.13, p = 0.004; Munro Cullum et al., 2014). Additionally, there was strong validity statistics (ICC ranged 0.77 -0.81; Marra et al., 2020). No studies found significant differences for means on HVLT-R Delayed Recall (Marra et al., 2020). Similar validity metrics were found for Delayed Recall as they were moderate (ICC ranged from 0.61 to 0.90; Marra et al., 2020). Furthermore, no information regarding standard deviation differences or Bland-Altman plots are discussed in this critical review.

WMS Logical Memory. As previously documented, Jacobsen et al. (2003) assessed reliability and equivalence for mode of administration on WMS Logical Memory I and II. Jacobsen et al. (2003) administered the test with no modifications in instructions for the VTC condition. There was a significant difference on WMS Logical Memory I for mode of administration means between FTF 15.1(3.75) and VTC 16.3(3.6) (p = 0.02; Jacobsen et al., 2003). However, there was no significant difference on WMS Logical Memory II for mode of administration means between FTF 13.6(3.8) and VTC 14.6(8.8) (p = .17; Jacobsen et al., 2003). FTF and VTC were found to be strongly correlated with each other as r > 0.80 for both WMS Logical Memory I and II (Jacobsen et al., 2003). Given the significant difference between means for WMS Logical Memory I, further investigation should be conducted to determine psychometric similarity and equivalence between mode of administration.

Benton Visual Retention Test. Jacobsen et al. (2003) conducted a study to assess reliability and equivalence for mode of administration on Benton Visual Retention Test (BVRT). Jacobsen et al. (2003) administered the test with no modifications in instructions for the VTC condition; items were transmitted via a document camera and the participants responded with the corresponding record booklets provided on the desk. There was no significant difference on BVRT correct response and error response for mode of administration means between FTF and VTC ($p = 0.27$; Jacobsen et al., 2003). FTF and VTC were found to be moderately correlated with each other as $r > 0.60$ for both BVRT correct response and error response (Jacobsen et al., 2003). Jacobsen et al. (2003) began to establish equivalence for mode of administration; however, further investigation needs to be conducted into the feasibility and psychometric similarity of the BVRT.

Eligibility Criteria

What Does the Neuropsychologist Need?

Similar to FTF, VTC neuropsychological assessment comes with its own standard procedures, technologies, and materials that need to be taken into consideration by a neuropsychologist when deciding if providing VTC neuropsychological assessments is a best fit for them. Studies done by Cullum et al. (2014), Wadsworth et al. (2014), Wadsworth et al. (2018), and Jacobsen et al. (2003) had specific remote testing rooms where typically the patient or participant was informed how to use the technology in the room and a direct Internet connection was made. Munro Cullum et al. (2014) had local staff present to help with VTC equipment if needed but they were not present during administration. Additionally, local staff introduced the equipment used during VTC that helped participants adapt (Munro Cullum et al., 2014). Munro Cullum et al. (2014), Wadsworth et al. (2014), Wadsworth et al. (2018), and Galusha-Glasscock et al. (2016) reported using a PC – Based Videoconferencing System (Polycomm iPower 680 Series) with 26” flat screen where the patient sat 30” away from the screen at a desk. Munro Cullum et al. (2014) and Jacobsen et al. (2003) suggest a bandwidth of at least 384 kbit as this was determined to be the minimum quantity needed for a synchronized/quality sound and picture. Jacobsen et al. (2003) used two videophones (Tanberg 5000) that were located in the psychologist’s office and a Polyspan view station. Picture and sound were transmitted via parallel ISDN units. Additionally, they used a document camera (JVC visual presenter AV-P700) to enhance resolution of visual and printed material. Jacobsen et al. (2003) suggest using a larger screen for the participants than the average computer size. The clinical review conducted by Marra et al. (2020) found only two studies that used personal laptops and a cloud-based video conferencing system; they reported there was insufficient evidence currently to use a cloud-based video conferencing platform. Marra et al. (2020) reported if a cloud-based video conferencing system is used, sufficiently fast and reliable Internet is required (>25mbit/s).

A report from Emory by Hewitt and Loring (2020) was the only article that discussed consent. Emory by Hewitt and Loring (2020) needed to obtain additional consent for telehealth neuropsychological assessment that was obtained verbally by asking birthdate, patient location, and asking the patient to agree to not record any part of the assessment. Patients were then informed of the procedures and how they differed from traditional procedures. If there were any concerns during diagnostic interview or the assessment the patient would be informed, this would be documented in their chart that in-person follow up would be best (Hewitt & Loring 2020). Emory's clinic provided staff with administration manuals that included information about using Zoom interfaces for assessment, how to adapt test instructions, and troubleshooting (Hewitt & Loring, 2020). Additionally, they practiced the ZOOM calls with mock patient where two others would also attend (Hewitt & Loring, 2020) with one person being the patient, one being the psychometrist, and the other observing (Hewitt & Loring, 2020). Hewitt and Loring (2020) sometimes would send out an initial email to the patient. The review indicated they also used a HIPAA compliant server that utilizes duo with a two-factor authorization to gain access to store patient responses in all test materials (Hewitt & Loring, 2020). Additionally, they engaged in a pre-assessment appointment where the psychometrists would inquire about a patient's eligibility for telehealth and would provide a Zoom training meeting (Hewitt & Loring, 2020). Additional information was provided regarding back up plans if technological issues were to arise. Lastly, they would utilize the screengrab feature in order to obtain pictures of client responses that would be then saved in their HIPAA compliant server (Hewitt & Loring, 2020).

What Populations are Best Suited for Televideo Assessments?

Demographics. Many of the above-mentioned studies specifically looked at certain population groups; most typically, older adults were utilized. A study conducted by Munro Cullum et al. (2014) used older adults with a mean age of 68.5 (46-90). Similar ages groups were utilized in studies conducted by Wadsworth et al. (2018; M = 66.10) and Wadsworth et al. (2016; M = 64.89). Montani et al. (1996) had the oldest age group with a mean age of 88 and Jacobsen et al. (2003) had the youngest age group with a mean age of 34.8. Munro Cullum et al. (2014), Wadsworth et al. (2014), Wadsworth et al.

(2018), Jacobsen et al. (2003) and Galusha-Glasscock et al. (2016) were primarily female participants. The vast majority of the studies only included White participants; however, Wadsworth et al. (2014) only included Native Americans in their study. Lastly, the majority of the participants had higher education typically with a mean education of 14 years (Galusha-Glasscock et al., 2016; Jacobsen et al., 2003; Munro Cullum et al., 2014; Wadsworth et al., 2018).

Cognition. The majority of the above-mentioned studies included both healthy controls and those that are cognitively impaired. Specifically, those included participants with a diagnosis of MCI and AD (Galusha-Glasscock et al., 2016; Montani et al., 1996; Munro Cullum et al., 2014; Wadsworth et al., 2014; Wadsworth et al., 2018). The study conducted by Jacobsen et al. (2003) only included healthy controls.

Exclusion Criteria. Unlike telephone neuropsychological assessments, VTC neuropsychological assessment did not include information regarding exclusion criteria for participants chosen in their studies period. Often in the telephone neuro psychological assessments, hearing difficulty was considered to be an exclusion criterion. Additionally, in computerized neuropsychological assessments comfort levels and ability to use technology was another exclusion or inclusion criteria. However, neither of these were addressed in the majority of the articles reviewed prior. In the review article submitted by Hewitt and Loring (2020), they provided a list of appropriate and inappropriate candidates specifically screening out both medico-legal cases and epilepsy surgery candidates because they felt these cases needed to be seen in person as they required comprehensive and widely validated assessment procedures. Candidates who lacked the ability to use technology were also considered inappropriate candidates. Finally, during the diagnostic interview, they used the MoCA as a screening tool and those with low scores would be recommended for follow up in person. Emory brings up the need for the provider to be able to trust the patient (Hewitt & Loring 2020). For example, if they say they did not hear something, the provider will need to believe them and provide that information again (Hewitt & Loring 2020).

Implications for Clinical Practice

Benefits of Televideo Assessments

Many articles have been published outlining the vast benefits of Telehealth services for medicine, and psychology; however, with the increasing research on neuropsychological assessment, there is limited research looking at the benefits of neuropsychological assessment. These articles do not outline benefit and did not review comfort levels with their participants. However, one can presume neuropsychological assessment provides many benefits to be able to assess and provide services to those who cannot typically access services. Specifically, Wadsworth et al. (2016) intentionally reviewed VTC neuropsychological assessment with Native Americans as they typically are an underserved population and providing remote services increases the likelihood of being able to access services. VTC neuropsychological assessments allow for a wider reach of the populations especially those individuals who have limited mobility and live in remote locations (Marra et al., 2020). Another review article conducted by Brearly et al. (2017) reported that participants found VTC neuropsychological assessments to be convenient and clinics indicated that VTC assessments reduced costs. Additionally, given the current COVID-19 pandemic, being able to provide neuropsychological assessment remotely increases safety for both patients and neuropsychologists. Being able to conduct a remote neuropsychological service to those patients who are currently quarantined increases efficiency in care.

Disadvantages of Televideo Assessments

Although there are many advantages to VTC neuropsychological assessment, currently there are many disadvantages to providing VTC neuropsychological assessment. Specifically, there is limited research with participants who have other neurological disorders besides cognitive decline such as Parkinson's Disease and Multiple Sclerosis. Considering that these can impact a person's abilities to pay attention, processing speed, and physical abilities, this may increase impact on their performance on VTC neuropsychological assessment. Wadsworth et al. (2016) was the only study that specifically utilized participants of color. As such, there is very limited research assessing the validity and equivalence of VTC neuropsychological assessments with participants who identify as people of color. Currently, ethnic minorities have a reduced likelihood to access medical care via remote services; as such, socioeconomic factors increase barriers to their ability to obtain remote telehealth services because ethnic minorities do not report owning personal computers at the same rate as Caucasian older adults (Perrin & Turner, 2019).

Another concern for VTC neuropsychological assessments arises around current normative data for traditional neuropsychological assessments being used for VTC neuropsychological assessments. For many of the studies reviewed, their aim in the research was to validate the test to be utilized for VTC neuropsychological assessments. The review article conducted by Brearly et al. (2017) found differences in means with small but significant effect sizes. Specifically, on timed test or when the presentation of test information can be impacted if disrupted, performance on VTC had a lower standard deviation and a small but significant effect size (*g* = -0.10; SE = 0.03; 95% CI [-0.16, -0.04], *p* <.001; Brearly et al., 2017). Additionally, there were main differences noted on animal categories, BNT, Clock draw, MOCA, HVLT-R, and Logical Memory I in the aforementioned studies. As such, further research needs to be conducted into the clinical significance of these main differences and research conducting on differences in standard deviations.

Unlike telephone neuropsychological assessments, VTC neuropsychological assessments often were longer in administration time and required clinicians or technicians to score in the moment while the patient was holding up their test material to the camera. Both Munro Cullum et al. (2014) and Wadsworth et al. (2018) indicated Longer administration time for VTC versus FTF.

Missing Pieces

Many of the aforementioned studies are an excellent start at providing validity in reliability indicators for the use of traditional neuropsychological assessments to be utilized in a VTC setting. However, there are many missing pieces that would increase many clinician's level of comfort with utilizing VTC neuropsychological assessments. Specifically, future research should begin to address mean differences and standard deviation differences. As noted previously in many of the studies, variance and standard deviation differences were not a focus in this study. Current normative data uses both means and standard deviations to derive cutoff scores, standard scores, SS scores, T scores, and Z scores. Additionally, many of the studies have short periods of time between test conditions and as such, this can inflate the ICC. Further research needs to be conducted to make sure inflated ICCs are not a result of short test/retest periods. Therefore, these leaves us with the question of, which normative data should be used?

Additionally, there is very limited research being conducted outside of a remote telehealth office where a participant attends testing sessions in a specific remote telehealth office. This leaves us with the question of how this research applies to traditional office settings where direct internet connections cannot be made and information regarding a person's internet speed and or office internet speed may be unknown. Furthermore, most of the research does not address what to do if there is poor internet connection or lagging because this can greatly affect both time tests and tests that require specific timed presentation of information. Lastly, what should be addressed in report writing for the clinical neuropsychologist if there are issues during testing and should it be noted that testing was done remotely?

Emerging Technology

Telehealth Clinics in Practice

In a study done by Harrel et al. (2014) at a telemedicine clinic associated with V-CAMP, the VHA telemedicine clinic was located in a major metropolitan medical center who severed three VA community Based Outpatient Clinic. This study looked at the outcome of 100 patients this clinic served with the addition of comprehensive neuropsychological assessments (Harrell et al., 2014). At the main location, a clinical neuropsychologist and a fellow were administering, interpreting, and providing feedback to these patients (Harrell et al., 2014). At the remote locations, telehealth clinical technicians were located and accompanied the patients in the evaluation room. These technicians helped coordinate teleconferencing and prepared stimuli for evaluations and would fax information following test completion (Harrell et al., 2014). These technicians had educational and occupational background including nursing and public administration (Harrell et al., 2014).

The procedures for test administration included folders that contained test materials and were organized by telehealth clinical technicians ordered in accordance with the administration predetermined by the neuropsychologist (Harrell et al., 2014). These folders were numbered so the patient could be instructed to remove and return materials from the corresponding folders during test administration (Harrell et al., 2014). The telehealth clinical technician was there to help manipulate the camera angle to correspond with different tests to allow the administrator to observe and provide feedback (Harrell et al., 2014). Of note, most patients were able to follow instructions without difficulty; however, telehealth clinical technicians were available as needed.

Harrell et al.'s (2014) clinic had a standard battery that included two assessments in the following domains: attention and concentration, language, visuospatial functioning, learning and memory measures both visual and verbal, and executive functioning, and psychological functioning and single measures of global cognitive functioning, psychomotor speed, and premorbid intellectual functioning. Administration time range from approximately 90 to 120 minutes and was divided into two testing sessions (Harrell et al., 2014). The technology utilizes two Cisco EX90 devices with 24" crystal display with a resolution of 1920 x 1200 (Harrell et al., 2014). The camera has a 1080p30 resolution with autofocus (Harrell et al., 2014). The VTC connection utilized VHA telehealth infrastructure capable of providing high speed digital connections with heavy encryption (Harrell et al., 2014).

The study conducted by Harrell et al. (2014) included patient outcomes. Of note, 87% of the first 31 patients referred for testing had an inaccurate neurocognitive diagnosis at the time of referral and VTC neuropsychological assessment was able help clarify and provide more accurate diagnosing (Harrell et al., 2014). There were high rates of patient acceptance in that all patients indicated being able to tolerate VTC neuropsychological assessment (Harrell et al., 2014). Additionally, no adverse outcomes attributable to VTC neuropsychological assessments were noted (Harrell et al., 2014). However, VTC was a trigger for paranoia for a patient with comorbid psychotic disorder (Harrell et al., 2014).

Clinical Pearls

Table 8 provides a summary of the above-mentioned research regarding validity, modifications, and if new normative data is need for each assessment reviewed. Table 8 can be used as a quick reference guide for clinicians when deciding what assessments to utilize during remote computerized assessments.

Table 8

Does Mode of Administration Matter for Televideo Neuropsychological Assessments

Test	Valid	Modifications	New normative data needed
MMSE	Yes	Yes	Only one study found mean differences generally FTF was comparable with VTC
MoCA	Yes	No	New cutoff scores may be need because mode of administration does impact psychometric similarity between FTF and VTC
RBANS	Yes	No	No differences in means
Digit Span	Variable	No	No differences in means
Seashore Rhythm Test	N/A	No	Yes, difference in means were found in one study
Groove Pegboard	Yes	No	No differences in means
Symbol Digit Modalities Test	Variable	Yes	No differences in means; however, no research with standard deviation difference or Bland-Altman Plots
Oral TMT	Yes	No	No differences in means; however, no research with standard deviation difference or Bland-Altman Plots
BNT	Yes	No	Variable, mean and standard deviation difference but low bias with the Bland-Altman Plots
WAIS Vocabulary	Yes	No	No differences in means; however, no research with standard deviation difference or Bland-Altman Plots
VOSP Silhouettes	Yes	No	No differences in means; however, no research with standard deviation difference or Bland-Altman Plots
Rey-O	N/A	Yes	No research was run on means and standard deviation differences
COWAT	Yes	No	Animal Category revealed mean difference but no information was provided regarding Bland-Altman Plots
Clock Draw	Yes	Yes	Variable, two studies found mean differences
HVLT-R	Yes	No	Variable, mean difference but low bias with the Bland-Altman Plots
Logical Memory	Yes	No	Yes, mean differences on Logical Memory I but none on Logical Memory II and no information was provided regarding standard deviation difference and Bland-Altman Plots
Visual Memory	Yes	No	No differences in means; however, no research with standard deviation difference or Bland-Altman Plots

*N/A= not applicable

CHAPTER V: DOES MODE OF ADMINISTRATION MATTER?

The previous sections explored various modes of administration for remote neuropsychological assessments; however, the question remains: Does mode of administration matter? An abundance of research was reviewed for telephone neuropsychological assessments, computerized neuropsychological assessments, and televideo neuropsychological assessments that provided arguments for the ease of transition from paper pencil tasks to remote testing via these modalities. The following section will provide clinicians with a better understanding of equivalence for mode of administration in research for each mode of administration and procedural guidelines if they are to engage in remote assessment.

Telephone Assessments

Many research articles were reviewed for telephone neuropsychological assessments. Very few of those articles ran analyses to determine mean and standard deviation differences. A main study reviewed found that many correlations where significant; however, a few were weaker than expected (digit span and category fluency, HVLT-R discrimination index). Given this information, there still needs to be further assessment on equivalence. Also, this study highlighted mean differences between mode of administration that shows the impact of telephone administration of normative data (Bunker et al., 2017). Additionally, modifications were often made to administer tests over the telephone. Specifically, Bunker et al. (2017) modified the BNT by having the interviewer read a short sentence describing an object and the participant was asked to name the object. However, clinically this significantly changes the purpose of the BNT and as such may account for the mean differences found. Given the current research, it appears that cognitive screeners such as the various versions of the TICS and the T-MoCA have the strong psychometric statistics with new cut off scores for telephone administration. For many of the standard neuropsychological assessments researched to be administered over the phone, limited analyses were run to determine equivalence and if new normative data was needed.

Clinical Implications

The intention of this clinical research project is to provide a procedural guideline to clinicians on best practice for the administration of remote neuropsychological assessments. Table 9 provides steps for the clinician to consider when providing telephone neuropsychological assessments. The American Psychological Association (2013), Bilder et al. (2020), and Grosch et al. (2011) provided guidelines in regards to televideo neuropsychological assessments; however, there are limited guidelines regarding conducting telephone neuropsychological assessments. Table 9 provides guidance and questions when a clinician is making the decision to utilize telephone neuropsychological assessments.

Table 9

Step-by-Step Guide for Telephone Neuropsychological Assessments

Steps	Instructions
1	Consult referral question Hewitt and Loring (2020) ruled out surgical and medical legal cases as such higher risk assessments may be best conducted in the office
2	Patient demographics – those that are hard of hearing or having higher anxiety may not be a best-fit for remote assessments
3	Caregiver support – some the assessments required caregiver support; therefore, discussing this with the caregiver will be important prior to conducting the assessment
4	Obtaining consent – reviewing the risks and benefits of telephone assessment that includes possible risks to confidentiality, impacts of modifications of standardized assessments, reduction of behavioral observations, and how this can impact the implications or conclusions gathered by the assessment.
5	Using Table 2 at the end of chapter 2 to help select the test list
6	Making sure to address modifications and implication of the assessment within the report.

Computerized Assessments

There is significant research in regards to the impact of computerized assessments when a traditional paper pencil assessment has been modified to be administered on the computer. For potential cognitive screener measures such as the CAMCI and dCDT, new normative data was developed. However, for many of the assessments that were transitioned into computerized assessments, the results were quite variable on whether there was equivalence between modes of administration. Specifically, the WCST, ToH, Stroop, and TMT had variability throughout a few studies with mean differences between modes of administration. Many of these articles found significant differences between means thereby indicating the need for new normative data when a traditional paper pencil task has been translated to a computerized assessment either via a computer or an iPad/tablet. Although there are many benefits that come with computerized assessments, further research needs to be conducted to ensure equivalence.

Clinical Implications

As mentioned previously, a primary goal of this clinical research project is to provide a procedural guideline on best practice for the administration of remote neuropsychological assessments for neuropsychologists. Table 10 provides steps for the clinician to consider when providing computerized neuropsychological assessments. Although the research review only evaluated computerized assessments that were directly developed from paper pencil tasks (and as such are still typically administered in person and are not sent to an individual to take on their own), there are still considerations for a neuropsychologist to contemplate when deciding to administer computerized assessments. The following guidelines were developed from the research reviewed and the American Psychological Association (2013), Bilder et al. (2020), and Grosch et al. (2011). Many research articles were reviewed for telephone neuropsychological assessments.

Table 10

Step-by-Step Guide for Computerized Neuropsychological Assessments

Steps	Instructions
1	Consult referral question to ensure the patient is the best candidate for computerized assessments
2	Patient demographics – it will be important to assess level of comfort/knowledge with computers as this may impact performance (Williams & Noyes, 2007); Possible TBI - Spreij et al. (2020) found that patients with a TBI had increased difficulty completing an iPad Assessment due to eyestrain and increased headache.
3	Technology – Cernich et al. (2007) reviewed the impact operating systems can have on the programs utilized to run the assessments; as such, it is very important that the operating system used matches what is reported in the research or manual. Additionally, when using iPads or tablets Spreij et al. (2020) discussed the need to make sure the patient is comfortable and knowledgeable on the system. Spreij et al. (2020) also found that some patients had difficulty with using a stylus; they found there was differences in feel and execution between traditional paper pencil and the iPad.
4	Obtaining consent – reviewing the risks and benefits of computerized assessment including the impacts of modifications from standardized assessments and the potential for a reduction in behavioral observations and how this can impact the implications or conclusions gathered by the assessment.
5	Using the Table 6 at the end of chapter 3 to help select the test list and identify which tests have the strongest psychometric evidence for use.
6	Making sure to address modifications and implication of the assessment within the report.

Televideo Assessments

Recently, there has been an influx of research conducted in regard to televideo neuropsychological assessments. Specifically given the COVID-19 pandemic, Marra et al. (2020) conducted a meta-analysis reviewing all of the current research on televideo neuropsychological assessments. This clinical research project not only reviewed the meta-analysis but also reviewed individual research articles on televideo neuropsychological assessments. Two main articles were reviewed by Munro Cullum et al. (2014), Grosch et al. (2015) and Jacobsen et al. (2003) found mode of administration impacted MoCA, Seashore Rhythm Test, Oral TMT, BNT, COWAT, Clock Draw, HVLT-R, and Logical Memory. Although many of these tests showed some mean differences, many of them were variable across different studies. However, Galusha-Glasscock et al. (2016) had strong psychometric properties for the RBANS. Additionally, there are some benefits to providing assessments via VTC; Jacobsen et al. (2003) found that participants felt they could focus better during VTC conditions. Although Hewitt and Loring (2020) did not conduct any research regarding equivalency, have been running a remote telehealth clinic utilizing personal computers and have found success in doing so.

Clinical Implications

A primary goal of this clinical research project is to provide a procedural guideline on best practice for the administration of remote neuropsychological assessments for neuropsychologists. Table 11 provides steps for the clinician to consider when providing VTC neuropsychological assessments. Although the research review only evaluated VTC assessments in telehealth clinics, Hewitt and Loring (2020) provided information regarding how their clinic conducted VTC neuropsychological assessments that were directly developed from paper pencil tasks (and as such are still typically administered in person and are not sent to an individual to take on their own), there are still considerations for a neuropsychologist to contemplate when deciding to administer computerized assessments. The following guidelines were developed from the research reviewed from the American Psychological Association (2013), Bilder et al. (2020), and Grosch et al. (2011)

Table 11

Step-by-Step Guide for Televideo Neuropsychological Assessments

Steps	Instructions
1	Consult referral question to ensure the patient is the best candidate for computerized assessments
2	Patient demographics – it will be important to assess level of comfort/knowledge with computers as this may impact performance (Williams & Noyes, 2007). There is limited research with Parkinson's Disease and MS. Munro Cullum et al. (2014) reviewed a battery of tests that targeted participants with dementia; primarily the research focused on older adults with normal cognition, MCI, and AD.
3	Technology and bandwidth – Hewitt and Loring (2020) utilized the medical version of ZOOM and APA (2013) provided a list of HIPPA compliant web-based video call systems. Additionally, Marra et al. (2020) reported if a cloud-based video conferencing system is used, sufficiently fast and reliable Internet is required (>25mbit/s). Hewitt and Loring (2020) screened patients' experience with web-based video conferencing systems and provided a tutorial to utilize the systems properly.
4	Obtaining consent – reviewing the risks and benefits of computerized assessment including the impacts of modifications from standardized assessments and the potential for a reduction in behavioral observations and how this can impact the implications or conclusions gathered by the assessment. Additionally, addressing the limitations in current research because no new normative data has been developed for VTC assessments.
5	Using the Table 8 at the end of chapter 4 to help select the test list and identify which tests have the strongest psychometric evidence for use.
6	Making sure to address modifications and implication of the assessment within the report.

Additional Considerations

Despite the wealth of research, the question remains to be where and how a neuropsychological begins to engage in remote neuropsychological assessments. In that many neuropsychologists do not have access to remote telemedicine clinics and given the current COVID-19 pandemic, what would be considered best practice to provide remote neuropsychological assessments. The following will provide a guideline based on the research reviewed and adapted from the American Psychological Association (2013), Bilder et al. (2020), and Grosch et al. (2011).

The first step would be to determine type of remote assessment best suits the neuropsychologist. Based on the abundance of research at this time, televideo assessments have the strongest statistical data to support use in clinical practice and the strongest psychometric evidence. Additionally, when making this decision, consider the referral question and if the research provided matches the neuropsychologist's intended assessment battery.

After the decision has been made regarding what mode of administration the neuropsychologists should also consider, what technology is needed and what does one do if the technology fails. What technology is needed? It will be imperative for the neuropsychologists to make sure that they have the adequate technology when providing remote neuropsychological assessments. Munro Cullum et al. (2014) and Jacobsen et al. (2003) suggest a bandwidth of at least 384 kbit, and Marra et al. (2020) suggested >25mbit/s for sufficiently fast and reliable internet. When providing VTC assessment, it will be important to assess the bandwidth of the patient's internet to ensure a quality video call. Given the vast majority of the research reviewed regarding televideo assessments were done in a telemedicine clinics, when conducting a televideo assessment via personal computer, it will be imperative for the neuropsychologists to have a plan in place if the Internet fails or if there is lag or buffering while conducting the assessment. None of the research reviewed discussed what to do if this were to happen. It will be important for the neuropsychologist to provide a plan prior to starting the assessment with the patient indicating what to do if the call were to be lost.

Additionally, another factor a neuropsychologist may consider when deciding which modality to use for remote neuropsychological assessment is the age of the patient and their level of comfort with technology. Williams and Noyes (2007) found differences in ability between novice and experienced computer users. It may be helpful to assess for a patient's level of computer experience because this may impact their performance. Hewitt and Loring (2020) assessed patients' computer experience and provided them with a tutorial prior to their VTC neuropsychological assessment to ensure the patient was able to utilize the web-based video conferencing system. Many studies addressed level of anxiety. It has been well established how anxiety can impact cognitive functioning; as such, it will be imperative for the clinician to assess the patient's level of anxiety with the use of technology and if their anxiety increases during the assessment due to technology use as this will need to be noted in behavioral observations.

An additional consideration for neuropsychologists is in regard to report writing in addressing modifications, deviations from standardization, and implications due to limitations of remote neuropsychological assessments. The American Psychological Association (2010) ethics code 9.02 specifically addresses this consideration. Specifically,

> 9.02(a) "Psychologists administer, adapt, score, interpret, or use assessment techniques, interviews, tests, or instruments in a manner and for purposes that are appropriate in light of the research on or evidence of the usefulness and proper application of the techniques. (b) Psychologists use assessment instruments whose validity and reliability have been established for use with members of the population tested. When such validity or reliability has not been established, psychologists describe the strengths and limitations of test results and interpretation." (APA, 2010, p. 1071)

Given this ethical code, it will be important for neuropsychologists to provide information within the report regarding limitations and modifications provided. Below is an example of how to address this with in the neuropsychological report.

> The neuropsychological assessment was conducted using remote telehealth methods (i.e., telephone, computerized, or televideo). This included modification of the standard administration procedures for typical face to face neuropsychological assessments. As such, there may be an impact of utilizing non-standardized assessment procedures because this has only partly been evaluated in current research. Although every effort was made to remain as standardized as possible, the implications arrived from this assessment such as the diagnostic conclusions and recommendations for treatment should be taken with caution.

Determining Equivalence

Although an abundance of research was reviewed, there was no consistent way to determine equivalence between modes of administration. Determining validity and reliability has been well established within the psychological community to provide adequate psychometric statistics on new and well-established neuropsychological assessment. However, when comparing modes of administration, no standard set of statistical analysis is utilized. Often correlation in regression statistical analysis is utilized; however, the difficulty with using these analyses that do not assess differences but assess the relationship between two variables; as such, they are not an adequate method to compare differences. The following section will propose a set of analyses to determine equivalence and suggest best practice to determine equivalence for future research.

As mentioned above, it will be important to continue to provide information regarding the new mode of administration's validity and reliability in comparison to traditional or the original assessment and as such, standard practice should continue for determining validity and reliability. Although a new mode of administration may be considered valid and reliable, this does not dictate information regarding normative data. Therefore, further analysis needs to be conducted to determine if the modes of administration are equivalent in regard to normative data. Two studies reviewed in this clinical research project utilized a combination of statistical analysis that are beneficial in determining equivalence between modes of administration.

An initial step in determining equivalence should include calculating both mean and standard deviation difference that can be done varies ways. Traditionally, comparing two measurements is often done by a paired samples T test analysis. Munro Cullum et al. (2014) utilized both the Bradley-Blackwood procedure and the Pitman test to examine biases between modes of administration for both means and standard deviations. Bartko (1994) reported that in statistics, the Bradley-Blackwood procedure analyzes differences for both means and variances at the same time whereas traditionally a paired sample T test analysis is utilized to assess for mean differences and a Pittman test is utilized to assess for variances or standard deviation differences (Bartko 1994). As such, Munro Cullum et al. (2014) followed up the Bradley-Blackwood procedure with a Pitman test when significant to determine if the significance was due to mean differences or standard deviation differences. Both procedures are needed to determine the source of the bias that will be imperative to discover if new normative data needs to be collected. Means and standard deviations are utilized in normative data in order to derive Z scores, T scores, and standard scores that are some of the typical ways to determine if impairment is present when comparing a person to the general population. As such, it is imperative to have proper normative data in order to indicate impairment any person's performance.

The Bland-Altman Plots were developed by Altman and Bland (1983) because these researchers found that correlations and linear regressions are unable to determine equivalence. Altman and Bland (1983) describe the Bland-Altman Plot as an analysis of differences in order to quantify agreement between two measures. This is done by finding statistical limits that are calculated using both mean and standard deviation differences between the two measurements (Altman & Bland, 1983; Bland & Altman, 1999). This will allow for the ability to find the limits of agreement; as such, the researchers would want the limits to be as close to zero as possible as zero that would indicate complete equivalence but, given standard of error and variability, this would likely be impossible (Altman & Bland, 1983). As such, Altman and Bland (1983) report that the smaller the difference the more agreement there is between the two measurements. Therefore, adding Bland-Altman Plots will provide another statistical analysis for determining equivalent and providing further evidence if two modes of administration are equivocal.

Lastly, an important part of neuropsychological assessments is the ability to determine impairment from the general population. When transforming a test to a new mode of administration, this may impact the test's ability to determine impairment. The study conducted by Spreij et al. (2020) utilized Lezak's distribution to determine if new normative data was needed. Specifically, Lezak's distribution indicates that less than 10%

of healthy controls or healthy patients should perform below the 10th percentile within the general population (Lezak et al., 2012). As such, this helps determine if the test can indicate impairment within a specific domain. Vermeent et al. (2020) conducted a study analyzing motive administration on the iPad for a battery of neuropsychological assessments and part of the analysis included a factor analysis to ensure that specific assessments loaded on the intended cognitive factor. Specifically, Vermeent et al. (2020) utilized the neuropsychological consensus model to determine the factor groups that included attention in working memory, processing speed, visual spatial processing, executive functioning, and memory. Language was not included because the researchers did not include language-based tests. This factor analysis allows for further understanding and confirmation that the mode of administration does not impact the intended use of the assessment.

Future Research

Remote administered neuropsychological assessments are of value because they allow for more frequent follow up as well as the ability to follow up with patients who are unable to attend their appointments physically in the office. However, further research needs to be conducted to determine equivalence. The vast majority of the research had limited demographics in regards to gender, race, and education. Specifically, Rapp et al. (2012) found differences in scores based on race; as such, further research should be conducted assessing the difference in race with mode of administration. Furthermore, Wadsworth et al. (2016) was the only study that specifically utilized participants of color. Additionally, in regards to remote assessment, a study conducted by Perrin and Turner (2019) found people of color own computers at a lower rate than White adults. The need to address race and socioeconomic status within neuropsychological assessment is widely supported. Furthermore, there is limited research on adults with lower levels of education. There is very limited research assessing the validity and equivalence of remote neuropsychological assessments with participants who have other neurological disorders beside MCI or AD.

Future research should also take into consideration technology experience and anxiety while using a technology in the impact this has on a patient performance. Specifically, Williams and Noyes (2007) was the only study that took into consideration technology experience; the researchers found differences between novice and experienced users.

The vast majority of the research conducted in regard to VTC neuropsychological assessments has been done in telehealth medical clinics. Therefore, research should be conducted comparing traditional neuropsychological assessments where the patient is at home and utilizing a web-based video conferencing system. As such, this will allow for information to be gathered regarding this method of administration and will provide a basis for clinicians who do not have access to this type of clinic. Additionally, this research will allow for a better understanding of the impact of an uncontrolled environment.

A large part of this clinical research project was to provide standardization to mode of administration with remote neuropsychological assessments. However, there is little standardization in how mode of administration is compared to determine equivalency. This clinical research project provided in an initial step at standardizing the way equivalency is conducted between modes of administration.

Finally, future research should take into consideration the need for new normative data especially on computerized assessments, telephone assessments, and televideo assessment. The goal of this clinical research project was to determine if mode of administration impacted normative data.

Summary

There are a multitude of benefits for remote neuropsychological assessments. Wadsworth et al. (2016) highlighted this benefit as they were able to assess populations that may not be able to attend traditional neuro psychological face to face assessments. Additionally, telephone assessment was frequently used to provide increased follow-up care with stroke patients. Computerized assessments also bring benefits with being able to assess reaction time with automated scoring that otherwise would be unable to be assessed in a traditional assessment. However, the purpose of this clinical research project was the determine if mode of administration mattered. Unfortunately, there is no straight answer because quite frequently the results of the impact of mode of administration varied across assessment and administration type. While there is a current wealth of research regarding mode of administration equivalency, the argument can be made that more information is needed regarding normative data for mode of administration in order to determine equivalency. Although using remote neuropsychological assessments does increase the population able to be assessed, there are still considerations that need to be taken to ensure accuracy in the assessment. The aim of this clinical research project was to determine equivalency in mode of administration; unfortunately, with the varied results of the data, there are some assessments that can be reliably administered remotely. More research needs to be conducted on the assessments with notable mean and standard deviation differences.

References

Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: The analysis of method comparison studies. *Journal of the Royal Statistical Society. Series D (The Statistician), 32*(3), 307-317. doi:10.2307/2987937

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. AERA.

American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct.* Washington, D.C.

American Psychological Association. (2013). *Guidelines for the practice of telepsychology.* Washington, D. C.

Aşkar, P., Altun, A., Cangöz, B., Cevik, V., Kaya, G., & Türksoy, H. (2012, April). A comparison of paper-and-pencil and computerized forms of Line Orientation and Enhanced Cued Recall Tests. *Psychological Reports, 110*(2): 383-96. doi: 10.2466/03.22.PR0.110.2.383-396. PMID: 22662393.

Bartko, J. J. (1994). Measures of agreement: A single procedure. *Statistics in Medicine*, *13*(5-7), 737–745. https://doi.org/10.1002/sim.4780130534

Benton, A. L., Varney, N. R., & Hamsher, K. D. (1978). Visuospatial judgment: A clinical test. *Archives of Neurology, 35*(6), 364–367. https://doi.org/10.1001/archneur.1978.00500300038006

Bilder, R. M., Postal, K. S., Barisa, M., Aase, D. M., Cullum, C. M., Gillaspy, S. R., … Woodhouse, J. (2020). Interorganizational practice committee recommendations/guidance for teleneuropsychology (TeleNP) in response to the COVID-19 pandemic. *The Clinical Neuropsychologist*, 1–21. https://doi.org/10.1080/13854046.2020.1767214

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*: 135-160.

Bracken, M. R., Mazur-Mosiewicz, A., & Glazek, K. (2018). Trail making test: Comparison of paper-and-pencil and electronic versions. *Applied Neuropsychology: Adult*, *26*(6), 522–532. https://doi.org/10.1080/23279095.2018.1460371

Brandt, J., Spencer, M., & Folstein, M. (1988). Telephone interview for cognitive status. *Cognitive and Behavioral Neurology*, *1*(2).

Brearly, T. W., Shura, R. D., Martindale, S. L., Lazowski, R. A., Luxton, D. D., Shenal, B. V., & Rowland, J. A. (2017). Neuropsychological test administration by videoconference: A systematic review and meta-analysis. *Neuropsychology review*, *27*(2), 174–186. https://doi.org/10.1007/s11065-017-9349-1

Brouillette, R. M., Foil, H., Fontenot, S., Correro, A., Allen, R., Martin, C. K., Bruce-Keller, A. J., & Keller, J. N. (2013). Feasibility, reliability, and validity of a smartphone-based application for the assessment of cognitive function in the elderly. *PloS One*, *8*(6), e65925.

Brooks, B. L., Strauss, E., Sherman, E. M. S., Iverson, G. L., & Slick, D. J. (2009). Developments in neuropsychological assessment: Refining psychometric and clinical interpretive methods. *Canadian Psychology/Psychologie Canadienne, 50*(3), 196–209. https://doi.org/10.1037/a0016066

Bunker, L., Hshieh, T. T., Wong, B., Schmitt, E. M., Travison, T., Yee, J., Palihnich, K., Metzger, E., Fong, T. G., & Inouye, S. K. (2017). The SAGES telephone neuropsychological battery: Correlation with in-person measures. *International Journal of Geriatric Psychiatry*, *32*(9), 991–999. https://doi.org/10.1002/gps.4558

Cernich, A. N., Brennana, D. M., Barker, L. M., & Bleiberg, J. (2007). Sources of error in computerized neuropsychological assessment. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *22 Suppl 1*, S39–S48. https://doi.org/10.1016/j.acn.2006.10.004

Chapman, J. E., Cadilhac, D. A., Gardner, B., Ponsford, J., Bhalla, R., & Stolwyk, R. J. (2019). Comparing face-to-face and videoconference completion of the Montreal Cognitive Assessment (MoCA) in community-based survivors of stroke. *Journal of Telemedicine and Telecare*, 1357633X19890788. Advance online publication. https://doi.org/10.1177/1357633X19890788

Dion, C., Arias, F., Amini, S., Davis, R., Penney, D., Libon, D. J., & Price, C. C. (2020). Cognitive correlates of digital clock drawing metrics in older adults with and without mild cognitive impairment. *Journal of Alzheimer's Disease: JAD*, *75*(1), 73–83. https://doi.org/10.3233/JAD-191089

Edwards, S., Brice, C., Craig, C., & Penri-Jones, R. (1996). Effects of caffeine, practice, and mode of presentation on Stroop task performance. *Pharmacology, Biochemistry, and Behavior*, *54*(2), 309–315.

Feldstein, S. N., Keller, F. R., Portman, R. E., Durham, R. L., Klebe, K. J., & Davis, H. P. (1999). A comparison of computerized and standard versions of the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, *13*(3), 303–313. https://doi.org/10.1076/clin.13.3.303.1744

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state:" A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Fong, T. G., Fearing, M. A., Jones, R. N., Shi, P., Marcantonio, E. R., Rudolph, J. L., Yang, F. M., Kiely, D. K., & Inouye, S. K. (2009). Telephone interview for cognitive status: Creating a crosswalk with the Mini-Mental State Examination. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *5*(6), 492–497. https://doi.org/10.1016/j.jalz.2009.02.007

Galusha-Glasscock, J. M., Horton, D. K., Weiner, M. F., & Cullum, C. M. (2016). Video teleconference administration of the repeatable battery for the assessment of neuropsychological status. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *31*(1), 8–11. https://doi.org/10.1093/arclin/acv058

Grosch, M. C., Gottlieb, M. C. & Cullum, C. M. (2011) Initial practice recommendations for teleneuropsychology. *The Clinical Neuropsychologist*, *25*:7, 1119-1133. doi: 10.1080/13854046.2011.609840

Grosch, M. C., Weiner, M. F., Hynan, L. S., Shore, J., & Cullum, C. M. (2015). Video teleconference-based neurocognitive screening in geropsychiatry. *Psychiatry Research*, *225*(3), 734–735. https://doi.org/10.1016/j.psychres.2014.12.040

Harrell, K. M., Wilkins, S. S., Connor, M. K., & Chodosh, J. (2014). Telemedicine and the evaluation of cognitive impairment: the additive value of neuropsychological assessment. *Journal of the American Medical Directors Association*, *15*(8), 600–606.

Hewitt, K., & Loring, D. (2020). Emory University telehealth neuropsychology development and implementation in response to the COVID-19 pandemic. *The Clinical Neuropsychologist*.

Jacobsen, S. E., Sprenger, T., Andersson, S., & Krogstad, J. M. (2003). Neuropsychological assessment and telemedicine: A preliminary study examining the reliability of neuropsychology services performed via telecommunication. *Journal of the International Neuropsychological Society*, *9*(3), 472–478. https://doi.org/10.1017/s1355617703930128

Lachman, M. E., Agrigoroaei, S., Tun, P. A., & Weaver, S. L. (2014). Monitoring cognitive functioning: psychometric properties of the brief test of adult cognition by telephone. *Assessment*, *21*(4), 404–417.

Lezak, M. D., Howieson, D.B., & Loring, D. W. (2012). *Neuropsychological assessment*. Oxford University Press.

Lipnicki, D. M., & Gunga, H. (2008). Physical inactivity and cognitive functioning: Results from bed-rest studies. *European Journal of Applied Physiology*, *105*(1), 27-35. doi:10.1007/s00421-008-0869-5

Lipnicki, D. M., Gunga, H., Belavý, D. L., & Felsenberg, D. (2009). Bed rest and cognition: Effects on executive functioning and reaction time. *Aviation, Space, and Environmental Medicine*, *80*(12), 1018-1024. doi:10.3357/asem.2581.2009

Marra, D. E., Hamlet, K. M., Bauer, R. M., & Bowers, D. (2020). Validity of teleneuropsychology for older adults in response to COVID-19: A systematic and critical review. *The Clinical Neuropsychologist*, *34*(7-8), 1411–1452. https://doi.org/10.1080/13854046.2020.1769192

Mataix-Cols, D., & Bartrés-Faz, D. (2002). Is the use of the wooden and computerized versions of the Tower of Hanoi puzzle equivalent? *Applied Neuropsychology*, *9*(2), 117–120. https://doi.org/10.1207/S15324826AN0902_8

Mather, D. N., Woodcock, R. W. (2001). Examiner Manual. *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside.

Montani, C., Billaud, N., Couturier, P., Fluchaire, I., Lemaire, R., Malterre, C., Lauvernay, N., Piquard, J. F., Frossard, M., & Franco, A. (1996). "Telepsychometry:" A remote psychometry consultation in clinical gerontology: Preliminary study. *Telemedicine Journal: The Official Journal of the American Telemedicine Association*, *2*(2), 145–150.

Monteiro, I. M., Boksay, I., Auer, S. R., Torossian, C., Sinaiko, E., & Reisberg, B. (1998). Reliability of routine clinical instruments for the assessment of Alzheimer's disease administered by telephone. *Journal of Geriatric Psychiatry and Neurology*, *11*(1), 18–24. https://doi.org/10.1177/089198879801100105

Moore, R. C., Swendsen, J., & Depp, C. A. (2017). Applications for self-administered mobile cognitive assessments in clinical research: A systematic review. *International Journal of Methods in Psychiatric Research*, *26*(4), e1562. https://doi.org/10.1002/mpr.1562

Munro Cullum, C., Hynan, L. S., Grosch, M., Parikh, M., & Weiner, M. F. (2014). Teleneuropsychology: evidence for video teleconference-based neuropsychological assessment. *Journal of the International Neuropsychological Society: JINS*, *20*(10), 1028–1033.

Mrazik, M., Millis, S., & Drane, D. L. (2010). The oral trail making test: Effects of age and concurrent validity. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *25*(3), 236–243. https://doi.org/10.1093/arclin/acq006

Noyes, J. M., & Garland, K. J. (2003). Solving the Tower of Hanoi: Does mode of presentation matter? *Computers in Human Behavior*, *19*(5), 579–592. https://doi.org/10.1016/s0747-5632(03)00002-5

Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, *51*(9), 1352–1375. https://doi.org/10.1080/00140130802170387

Parsey, C. M., & Schmitter-Edgecombe, M. (2013). Applications of technology in neuropsychological assessment. *The Clinical Neuropsychologist*, *27*(8), 1328–1361.

Pendlebury, S. T., Welch, S. J., Cuthbertson, F. C., Mariz, J., Mehta, Z., & Rothwell, P. M. (2013). Telephone assessment of cognition after transient ischemic attack and stroke. *Stroke*, *44*(1), 227–229. https://doi.org/10.1161/strokeaha.112.673384

Perrin, A., & Turner, E. (2020, August 27). *Smartphones help Blacks, Hispanics bridge some – but not all – digital gaps with Whites*. https://www.pewresearch.org/fact-tank/2019/08/20/smartphones-help-blacks-hispanics-bridge-some-but-not-all-digital-gaps-with-whites/.

Rankin, M. W., Clemons, T. E., & Mcbee, W. L. (2005). Correlation analysis of the in-clinic and telephone batteries from the AREDS cognitive function ancillary study. AREDS Report No. 15. *Ophthalmic Epidemiology*, *12*(4), 271–277. https://doi.org/10.1080/09286580591003815

Rapp, S. R., Legault, C., Espeland, M. A., Resnick, S. M., Hogan, P. E., Coker, L. H., … Shumaker, S. A. (2012). Validation of a cognitive assessment battery administered over the telephone. *Journal of the American Geriatrics Society*, *60*(9), 1616–1623. https://doi.org/10.1111/j.1532-5415.2012.04111.x

Roccaforte, W. H., Burke, W. J., Bayer, B. L., & Wengel, S. P. (1994). Reliability and validity of the short portable mental status questionnaire administered by telephone. *Journal of Geriatric Psychiatry and Neurology*, 7(1), 33–38. https://doi.org/10.1177/089198879400700107

Salnaitis, C. L., Baker, C. A., Holland, J., & Welsh, M. (2011). Differentiating Tower of Hanoi performance: interactive effects of psychopathic tendencies, impulsive response styles, and modality. *Applied Neuropsychology*, *18*(1), 37–46. https://doi.org/10.1080/09084282.2010.523381

Saxton, J., Morrow, L., Eschman, A., Archer, G., Luther, J., & Zuccolotto, A. (2009). Computer assessment of mild cognitive impairment. *Postgraduate Medicine*, *121*(2), 177–185.

Shum, D., Murray, R. A., & Eadie, K. (1997). Effect of speed of presentation on administration of the logical memory subtest of the Wechsler memory scale-revised. *The Clinical Neuropsychologist, 11*, 188–191.

So, C., & Pierluissi, E. (2012). Attitudes and expectations regarding exercise in the hospital of hospitalized older adults: A qualitative study. *Journal of the American Geriatrics Society, 60*(4), 713-718. Retrieved from

Spreij, L. A., Gosselt, I. K., Visser-Meily, J., & Nijboer, T. (2020). Digital neuropsychological assessment: Feasibility and applicability in patients with acquired brain injury. *Journal of Clinical and Experimental Neuropsychology*, *42*(8), 781–793. https://doi.org/10.1080/13803395.2020.1808595

Steinmetz, J. P., Brunner, M., Loarer, E., & Houssemand, C. (2010). Incomplete psychometric equivalence of scores obtained on the manual and the computer version of the Wisconsin Card Sorting Test? *Psychological Assessment*, *22*(1), 199–202. https://doi.org/10.1037/a0017661

Tun, P. A., & Lachman, M. E. (2006). Telephone assessment of cognitive function in adulthood: The brief test of adult cognition by telephone. *Age and Ageing*, *35*(6), 629–632. https://doi.org/10.1093/ageing/afl095

Vermeent, S., Dotsch, R., Schmand, B., Klaming, L., Miller, J. B., & van Elswijk, G. (2020). Evidence of validity for a newly developed digital cognitive test battery. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00770

Wadsworth, H. E., Dhima, K., Womack, K. B., Hart, J., Jr, Weiner, M. F., Hynan, L. S., & Cullum, C. M. (2018). Validity of teleneuropsychological assessment in older patients with cognitive disorders. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *33*(8), 1040–1045. https://doi.org/10.1093/arclin/acx140

Wadsworth, H. E., Galusha-Glasscock, J. M., Womack, K. B., Quiceno, M., Weiner, M. F., Hynan, L. S., Shore, J., & Cullum, C. M. (2016). Remote neuropsychological assessment in rural American Indians with and without cognitive impairment. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *31*(5), 420–425. https://doi.org/10.1093/arclin/acw030

Wagner, G. P., & Trentini, C. M. (2009). Assessing executive functions in older adults: A comparison between the manual and the computer-based versions of the Wisconsin Card Sorting Test. *Psychology & Neuroscience*, *2*(2), 195–198. https://doi.org/10.3922/j.psns.2009.2.011

Wechsler, D. (2008). *Wechsler adult intelligence scale): Technical and interpretive manual* (4th ed.). San Antonio, TX: Pearson.

Wechsler, D. (2009). *Wechsler memory scale: Technical and interpretive manual* (4th ed.). San Antonio, TX: Pearson.

Williams, D. J., & Noyes, J. M. (2007). Effect of experience and mode of presentation on problem solving. *Computers in Human Behavior, 23*(1), 258–274. https://doi.org/10.1016/j.chb.2004.10.011

Wilson, R. S., Leurgans, S. E., Foroud, T. M., Sweet, R. A., Graff-Radford, N., Mayeux, R., Bennett, D. A., & National Institute on Aging Late-Onset Alzheimer's Disease Family Study Group (2010). Telephone assessment of cognitive function in the late-onset Alzheimer's disease family study. *Archives of Neurology*, *67*(7), 855–861.

Wong, A., Nyenhuis, D., Black, S. E., Law, L. S., Lo, E. S., Kwan, P. W., Au, L., Chan, A. Y., Wong, L. K., Nasreddine, Z., & Mok, V. (2015). Montreal Cognitive Assessment 5-minute protocol is a brief, valid, reliable, and feasible cognitive screen for telephone administration. *Stroke*, *46*(4), 1059–1064. https://doi.org/10.1161/STROKEAHA.114.007253

Zygouris, S., & Tsolaki, M. (2015). Computerized cognitive testing for older adults: A review. *American Journal of Alzheimer's Disease and Other Dementias*, *30*(1), 13–28.