UTILIZING PERFORMANCE VALIDITY TESTING AND SOCIAL COGNITION MEASURES FOR TELENEUROPSYCHOLOGICAL ASSESSMENT

A Dissertation by

Shannon Bowerman Lavigne

Master of Arts, Wichita State University, 2018

Bachelor of Science, University of Central Missouri, 2015

Submitted to the Department of Psychology
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

July 2022

UTILIZING PERFORMANCE VALIDITY TESTING AND SOCIAL COGNITION MEASURES FOR TELENEUROPSYCHOLOGICAL ASSESSMENT

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy with a major in Psychology.

C. Brendan Clark, Committee Chair

Rachal A. Petts, Committee Member

Samantha Slade, Committee Member

Ryan W. Schroeder, Committee Member

Michael Birzer, Committee Member

Accepted for the College of Liberal Arts and Sciences

Andrew Hippisley, Dean

Accepted for the Graduate School

Coleen R. Pugh, Dean

ACKNOWLEDGMENTS

This dissertation is dedicated the numerous individuals (and animals) that supported and believed in me despite my own doubts through the process. I am here because of all of you, and I hope I can continue to make you proud.

To Dr. Deac Dorr for graciously accepting me into your laboratory despite knowing you were going to be retiring; you gave me the opportunity that opened the door to where I am now. To Dr. Brendan Clark, thank you for welcoming me into your lab when I was a homeless graduate student; the support, guidance, and encouragement you have offered is insurmountable and I cannot thank you enough for pushing me to chase my dream of neuropsychology.

To my mentors and supervisors at KU, Drs. Ryan Schroeder, Phillip Martin, and Kelli Netson-Amore, thank you for everything. My passion to understand the brain and behavior relationship was cultivated within your office walls and the classes you offered. It was through your guidance that I was able to find what I truly loved in psychology. Your example has proven that neuropsychology is a career to be proud of, one that can be fun (albeit nerdy), and one I know I will look forward to every day. Thank you for the many years of training and supervision that has catapulted me to my future.

To my parents, Reggie and Lesa. Since I was a young child, you both have instilled a belief that I could do anything I put my mind to. The emphasis you put on education and helping others in this world created the foundation that brought me here. To my brother, Sean, thank you for always answering my late-night calls and sending the kindest messages when I needed them most. You have always been one of my biggest cheerleaders and I cannot express how grateful I am to be your sister.

To my spouse, Mark, and our three pets. All four of you have been the driving force behind-the-scenes through this journey. The pep talks, the cuddles from the bubs, and the enormous amount of love in our home made the long hours and immeasurable amount of stress worth it. I am so excited to get to spend time together, without graduate school hanging over our heads. Thank you for your support amongst the many celebrations and the frequent rejections. Cheers to many more years (and maybe some more degrees).

Finally, to my support system of friends. Some of my friendships - Natalie, Sam, and Liz - have spanned numerous decades, countries, and life events. The three of you believed in me from the beginning. Thank you for grounding me and reminding me of my worth outside of graduate school. It was the three of you that always reminded me to follow my heart and my dreams, to not settle for anything less than what was the best for me. To the members of my cohort - Heidi and Charles - and the crew that helped me survive - Hilary, Karen, Jamie, and Ashley - this group has been the most supportive lifeline throughout this process. I am so lucky to have found all of you in the hecticness that is graduate school. To all of my friends, I know all of you will continue on to do amazing things. I am so excited that our world has you all in it and our communities are better because of each and every one of you.

Whether it has been since birth or since the start of graduate school, all of you have shaped my life in a positive way and there is no way I could adequately express my gratitude in this acknowledgement. Thank you all. From the bottom of my heart.

ABSTRACT

The purpose of this current project was to determine the feasibility of performance validity testing and social cognition measures using videoconferencing. The literature concerning teleneuropsychological assessment is still relative bare at this point when compared to telehealth utilization in other healthcare fields. With the COVID-19 pandemic, this gap became apparent as clinicians around the world attempted to piece together batteries that would support their clinical diagnoses while maintaining standardization through videoconferencing. No studies to date had evaluated the feasibility of administering performance validity and social cognition measures by means of videoconferencing. This study selected a core of cognitive, performance validity, and social cognition measures that were administered to participants over videoconferencing while they remained in their homes. Scores were compared to comparison groups from the same Midwestern University that were administered the same measures through face-to-face administration collected for previous studies. The results of this study found that numerous cognitive and performance validity measures did not produce comparable results across testing modalities. There were no significant differences observed on a measure of social cognition. Furthermore, the results of this study suggested continued research is needed to evaluate the effectiveness of administering assessments over videoconferencing and the feasibility of having participants remain in their homes during these evaluations.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

The concept of modern telemedicine emerged as early as 1879 when an article in *The Lancet* documented an unidentified family contacting their physician by telephone regarding their child's potential croup. The family reportedly lifted the child to the phone and the doctor listened to the child's cough, ultimately assuring the family that the child did not have croup (Aronson, 1977). As telecommunications advanced, technology-assisted diagnosis expanded across numerous practices. The first use of video communications for telepsychiatry evaluations is often dated to 1959 when the University of Nebraska developed a system where neurological examinations were video broadcast to medical students (Wittson & Benschoter, 1972). This same program cultivated a two-way, televised system with a closed circuit between a state hospital and the university separated by 112 miles to assist with consultations and interventions in their psychiatry and neurology departments (Institute of Medicine, 1996; Perednia & Allen, 1995). This program emphasized the importance of positioning the camera so that nonverbal cues like facial expressions and gestures could be observed by practitioners (Wittson, Affleck, & Johnson, 1961). When participants of these programs were assessed as to their willingness and concern related to receiving group therapy through a videoconferencing format virtually all of them reported minimal concern (Wittson et al., 1961). With the overarching success of this program, other programs began implementing telepsychology practices to further capitalize on telemedicine's advantages (Turvey & Myers, 2012).

With developing technologies and awareness of increased need, the United States (U.S.) federal government created the Office for the Advancement of Telehealth in 1987 (a division created within the Federal Office of Rural Healthcare; Health Resources & Services

Administration, 2019). The U.S. government helped fund numerous programs using grant monies to assist rural communities facing health disparities due to their distance from specialized healthcare facilities (Health Resources & Services Administration, 2019). Telemedicine programs have continued to expand over the past 50 years, and today approximately 76 percent of hospitals utilize “videoconferencing, remote monitoring, electronic consults, and wireless communications” to deliver a portion of their patient care (American Hospital Association, 2019). As medical systems have incorporated telecommunication systems to provide care, mental health providers have sought to utilize the same modalities of treatment.

Given telehealth’s expanding role in providing healthcare services, a joint task force was created in 2011 with members from the American Psychological Association, Association of State and Provincial Psychology Boards (ASPPB), and American Psychological Association Insurance Trust (APAIT) (American Psychological Association Practice Organization, 2014). This joint task force published the guidelines regarding the practice of *telepsychology* in 2013 after they were approved by the American Psychological Association Council of Representatives. This organization defined “telepsychology” as “the provision of psychological services using telecommunication technologies” (American Psychological Association, 2013, p. 791). They further articulated telecommunication technologies “include but are not limited to telephone, mobile devices, interactive videoconferencing, e-mail, chat, text, and Internet” (American Psychological Association, 2013, p. 792). A national survey of psychologists in 2018 found approximately 42.5 percent of practitioners employed some form of online delivered modalities to provide treatment to patients (Glueckauf et al., 2018). Often services are used for therapy or intake services, but there is a growing need for assessment and evaluation services to be expanded to populations that are not readily able to visit clinics. The American Psychological

Association recognized this need and further articulated telecommunication guidelines addressing the need and concerns related to telepsychological assessment.

One of the guidelines dictated by the American Psychological Association is geared towards testing and assessment practices through the use of telecommunication services. The guideline further discusses the importance of maintaining the "integrity of the psychometric properties of the test or assessment procedure" (2013, p. 798). The use of testing and assessment is paramount in assisting clinicians with diagnosis and treatment recommendations. A national survey in 2017 reported 84.8 percent of clinicians endorsed use of assessment to assist with diagnosis, 79.2 percent administer testing for treatment recommendations, and 70.6 percent provide measures to screen and document any potential cognitive or neuropsychological deficits (Wright et al., 2017). In particular, the field of neuropsychology relies heavily on the use of psychometric assessments as a recent survey suggests neuropsychologists spend just under 60 percent of their work hours utilizing assessment (Rabin, Paolillo, & Barr, 2016).

Consistently, testing has primarily been administered through the use of pen and paper instruments alongside an examiner. The emergence of computerized testing in the field of neuropsychology began in the 1980s (Kane & Kay, 1992). Even so, after three decades, an emerging field of teleneuropsychology has developed with 17.7 percent of neuropsychologists using video conferencing for intakes, assessment, and feedback for a variety of patients that cannot travel to a clinic, such as rural populations, elderly individuals, and incarcerated persons (Stolwyk & Hammers, 2020). Despite this seemingly large percentage of individuals utilizing these services, only 6.7 percent of neuropsychologists report using videoconferencing services for actual assessment purposes (Stolwyk & Hammers, 2020). In part, this is due to the limited research surrounding teleneuropsychology and obvious concern as related to the validity and reliability of testing instruments administered through a videoconferencing system. As mentioned above, one of the greatest concerns regarding the use of telepsychology for assessment is the concern related to maintaining the integrity of the instruments. A variety of cognitive measures have been found to be valid and reliable, with no significant differences between testing completed face-to-face or through videoconferencing (Brearly et al., 2017; Marra et al., 2020a). Additionally, many testing publishing companies have produced computerized self-report measures. There appear to be no publications that have examined performance validity or social cognition measures administered through a videoconferencing system.

Research and use of telemedicine or telehealth has grown exponentially over the last few decades, with an increased focus during the 2020 COVID-19 pandemic as the need for accessible treatment has been at the forefront in healthcare interventions. Preliminary insurance claim tracking has found that healthcare interventions through telemedicine or telepsychology services have increased by 5679.57 percent when comparing May 2019 to May 2020 (American Medical Association, 2020). The purpose of this project was to assist with filling the omission in the current literature by assessing the validity of performance validity and social cognition assessments when administered through a videoconferencing format. In efforts to ensure participant and examiner safety, all testing was completed remotely, utilizing the practice of teleneuropsychology. Results were compared to comparison groups previously gathered utilizing standardized face-to-face administration. Findings from this study could provide further information to assist with expanding current teleneuropsychological assessment. This in turn could facilitate a widening of available services to communities where health disparities are largest.

CHAPTER TWO

LITERATURE REVIEW

History of Technology in Psychiatric Services

Telehealth practices have been intermittently used to supplement physical and mental healthcare for over sixty years (Benschoter Wittson, & Ingham, 1965; Maxmen, 1978; Straker, Mostyn, & Marshall, 1976; Nickelson, 1998). One of the earliest psychiatric uses was implemented by the Nebraska Psychiatric Institute, a department of the University of Nebraska School of Medicine, in 1959 (Benschoter, 1967). Upon receiving a seven-year grant from the National Institute of Mental Health, a two-way closed-circuit television system was established spanning the distance of 112 miles, connecting the medical school and the Norfolk State Hospital (Nickelson, 1998; Perednia & Allen, 1995). The closed-circuit system allowed psychiatric and neurological consultations and psychological interventions to be completed through an electronic “face-to-face” format (Benschoter, Eaton, & Smith, 1965; Wittson et al., 1961). The program also hosted numerous grand rounds and lecturer series spanning the two facilities (Benschoter et al., 1965; Wittson & Dutton, 1957). One of the reasons for establishing this connection was to mitigate attrition rates at the hospital, in addition to providing medical students opportunities for furthering their education (Benschoter, 1967). Records showed implementing this closed-circuit connection allowed for the Norfolk State Hospital to discharge 47 percent of its population within three years in part due to increasing the number of staff able to provide services to their patients (Wittson & Benschoter, 1972).

Similar practices were established across various programs by 1970 through federal funding, including Massachusetts General Hospital, Dartmouth Medical School, Claremont General Hospital, Mount Sinai School of Medicine, Logan International Airport, and numerous

Veterans Hospitals (Maxmen, 1978; Solow, Weiss, Bergen, & Sanborn, 1971). By late 1973, thousands of individuals were seen for consultations, evaluations, psychopharmaceutical prescriptions, or therapy throughout the programs (Dwyer, 1973; Murphy & Bird, 1974). Clients included "children, teachers, prisoners, probation officers, and youth agencies" (Maxmen, 1978, p. 451). Other programs expanded to bring treatments to rural populations (Solow et al., 1971). With the growth of these technologies, correctional facilities and government programs were centers for research when assessing telehealth's feasibility and functionality.

In 1971, Deer Island House of Correction supplied their residents with monthly psychiatric evaluations through "interactive television" (videoconferencing) which allowed a significant increase in the time allotted for offenders with healthcare professionals (Dwyer, 1973). The U.S. government also expanded programs designed to assess and evaluate military personnel and astronauts while in space (Darkins & Gary, 2000). One interesting program (Space Technology Applied to Rural Papago Advanced Health Care; STARPAHC) targeted astronauts and residents of an isolated Tohono O'odham Indigenous Peoples reservation using satellite communications to provide medical services (Institute of Medicine, 1996). Despite programs like the ones described above, grant funding had started to run out. Given the costly cutting-edge technology these services required, most of these systems were forced to forego their telehealth services (Bashshur, 1997; Dwyer, 1973; Maxmen, 1978). Per Perednia and Allen (1995), only one formal telehealth program that was initiated prior to 1986 maintained its practice into the 1990s.

By the mid-1990s, the U.S. government as well as several state governments had begun offering initiatives to promote telemedicine. Monetary support was provided by tax dollars, regulatory judgments, and federal monies allotted to the states (Lipson & Henderson, 1996).

Despite financial constraints, correctional, government, and military sectors continued to find ways to utilize telehealth protocols. Efforts to promote telemedicine emphasized its benefit to supporting rural communities, offering necessary interventions and treatments for incarcerated populations, and further addressing the shortage of medical and psychiatric providers when compared to the number of individuals needing assistance (Brecht, Gray, Peterson, & Youngblood, 1996).

Telemedicine continued to expand, allowing the field of telepsychology to grow in a separate direction from that of telepsychiatry. Telepsychology is defined as "the provision of psychological services using telecommunication technologies…technologies include but are not limited to telephone, mobile devices, interactive videoconferencing, email, chat, text, and internet" (American Psychological Association, 2013, p. 791). Early telepsychology studies utilized videoconferencing or telephone services to provide therapeutic interventions (Lovell et al., 2006; Mohr et al., 2000; Nelson, Barnard, & Cain, 2003; Swinson, Fergus, Cox, & Wickwire, 1995). Even so, early meta-analyses showed the literature remained small with only two studies comparing the efficacy of teleinterventions to face-to-face psychotherapy (Bee et al., 2008).

In more recent times, the use of teleservices has grown exponentially. Prior to the outbreak of COVID-19 in 2020, only two percent of psychologists reported using telepsychology services in a survey completed in 2000; there was a minimal difference in providers surveyed in 2019 (Vanden Bos & Williams, 2000; Piece, Perrin, & McDonald, 2020). With that said, it was anticipated that telepsychology would expand with developing mobile devices, cloud computing, virtual reality, and videoconferencing. With onset of COVID-19, telepsychology has seen a growth of services and providers utilizing its practices (from 29 percent of providers using some form of telepsychology occasionally to 83 percent of psychologists using it exclusively after the onset of COVID-19; Sammons, VandenBos, & Martin, 2020). When assessing its efficacy, psychotherapy provided through telecommunication services has shown to be effective in treating a variety of disorders. These include anxiety disorders (Rees & Maclaine, 2015; Yuen et al., 2013), autism spectrum disorders (Ferguson, Craig, & Dounavi, 2019; Hepburn, Blakeley-Smith, Wolff, & Reaven, 2016), bipolar disorders (Hidalgo-Mazzei et al., 2015), chronic pain (Connolly et al., 2018; Glynn et al, 2020), family and couples therapy (Wrape & McGinn, 2018), depression (Kim, Gellis, Bradway, & Kenaley, 2019; Lichstein et al., 2013; Massoudi et al., 2019), eating disorders (Shingleton, Richards, & Thompson-Brenner, 2013), insomnia (Holmqvist, Vincent, & Walsh, 2014), obsessive-compulsive disorder (Brand & McKay, 2013; Himle et al., 2006), posttraumatic stress disorder (Bolton & Dorstyn, 2015; Glassman et al., 2017; Sloan et al., 2011), psychotic disorders (Lawes-Wickwar, McBain, & Mulligan, 2018), suicidality and self-harm behaviors (Fairchild, Ferng-Kuo, Rahmouni, & Hardesty, 2020; Gros et al., 2011), and substance use disorders (Lin et al., 2019; Young, 2011).

Computerized Testing

As traditional psychology has embraced technological advances, so did the subspecialty of neuropsychology. Neuropsychology is a branch of psychology which is centered on using cognitive assessments, with goals of understanding how changes in brain functioning impact cognition. Although the use of assessment has been around for decades, the shift towards integrating technology and assessment occurred relatively recently. Initially, the Department of Defense set out to develop computerized assessments that measured how a person’s functioning and performance is impacted by conditional stressors (e.g., lack of sleep, temperature, hunger) or drugs (Kane & Kay, 1992). This work led to the creation of the Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB; Englund et al., 1987; Schlegel, Gilliland, &

Crabtree, 1992). The UTC-PAB, which consisted of 25 separate tasks administered on a computer, allowed military personnel to assess the effects certain administered drugs (e.g., carbon monoxide) and conditions (e.g., noise or heat) had on individuals when they were asked to perform divided attention or highly difficult tasks (Reeves, Thorne, Winter, & Hegge, 1989; Perez, Masline, Ramsey, & Urban, 1987). In the community, other tests began to be administered through an automated process, such as the Raven's Progressive Matrices, which used a slide projector to change the stimuli (Waterfall, 1979). Building on the technological developments of singular assessments, larger entities, such as military and government agencies, began composing entire computerized batteries.

The military has continued to use automated and computerized testing with the developments of the Immediate Post-Concussion Assessment and Cognitive Test (ImPACT; Lovell, 2007; Resch et al., 2013), CogScreen (Kay, 1995), Defense Automated Neurobehavioral Assessment (DANA; Lathan et al., 2013), Spaceflight Cognitive Assessment Tool for Windows (WinSCAT; Kane, Short, Spies, & Flynn, 2005) and Automated Neuropsychological Assessment Metrics Military Battery (ANAM; Proctor et al., 2017; Reeves, Winter, Bleiberg, & Kane, 2007). Other assessments and test publishers have shifted their administration modality from pencil and paper to computerized testing. Examples of this include the Halstead Category Test (Reitan & Wolfson, 1985), Wisconsin Card Sorting Test (Heaton, 2003), continuous performance tests (Conners CPT-3; Multi-Health Systems, 2014), and Cognivue (Cahn-Hidalgo, Estes, & Benabou, 2020), a new computerized battery of multiple neuropsychological assessments. Again capitalizing on technological innovations, large assessment publishing companies produced interactive online systems allowing a selection of their assessments to be available for online administration in recent years. For example, Pearson deployed their Q-Interactive system, which allows most of their tests to be administered on a tablet (2012), Psychological Assessment Resources created PARiConnect (2013), and Multi-Health Systems published MHS Online (2009). These seemed like reasonable means to further expand the reach of psychological assessment considering computers and handheld mobile devices were becoming the norm for the average citizen.

Despite the increased establishment of computerized assessment, this movement provided little additional assistance in reaching rural or at-risk populations with limited households owning technology to accompany this growth. With advancing technologies, the number of U.S. households with access to a computer has increased drastically over the last three decades as only 8.2 percent of individuals had access to a household computer in 1984 and 52 percent of homes in 2000 (File & Ryan, 2013). Whereas, census data published in 2017 suggested 90.8 percent of homes had a computer or smartphone, and 83.9 percent of households had access to the internet (U.S. Department of Commerce, 2017). Even with the increase in accessible technology, providers maintain a seemingly archaic use of pencil and paper test protocols. A survey conducted by Rabin et al., (2014) found that when neuropsychologists were asked about what instruments they use in their clinics, only six percent of assessments reported were computerized. Additionally, providers indicated they "rarely" use automated assessments. Although computerized and automated assessment technologies have been present for many years, the tests have not been designed for remote or internet-based administration. The realms of technology in neuropsychology and videoconferencing for intervention purposes have been present for decades, but the integration of these practices is relatively recent. Given the pressing COVID-19 pandemic, it is imperative that neuropsychology catch up with various medical practices to ensure that patients are receiving the care they need in a timely fashion.

Teleneuropsychology

Teleneuropsychology is defined as “the application of audiovisual technologies to enable remote clinical encounters with patients to conduct neuropsychological assessments” (Bilder et al., 2020, pg. 2). Per a search of PubMed, PsycInfo, and PsycArticles, the first mention of the term “teleneuropsychology” occurred in 2011 (Grosch, Gottlieb, & Cullum). Despite this, there have been numerous studies utilizing telecommunications and videoconferencing for neuropsychological services prior to this date. In the mid-1980s, Kent & Plomin (1987) established a battery of spatial and analogy tests that could be administered over the phone; correlations between tele-assessment and face-to-face administration were determined to be 0.65 with test-retest reliability ranging from 0.70-.95. This concept was further investigated by Ball and McLaren (1997) when they discussed the use of telephones for cognitive assessment. Additional studies assessed intelligence, memory, verbal fluency, and screened for cognitive status through the use of a telephone (Nesselroade et al., 1988; Roccaforte et al., 1992; Desmond, Tatemich, & Hanzawa, 1994; Plassman et al., 1994; Gatz et al., 1995; Menon et al., 2001). The use of the telephone within cognitive assessment marks the first integration of technology in the field of neuropsychology. This was further expanded as innovations continued and videoconferencing became readily available.

Teleneuropsychological services through the use of videoconferencing initially focused on clinical interviews and provided feedback to patients and their families. One of the seemingly instrumental first-uses of “interactive video conferencing” for neuropsychological services occurred in Kansas in 1993 (Tröster et al., 1995). In an attempt to reach rural populations in Kansas that needed neuropsychological evaluation, a psychometrician was trained and employed at an outreach center in Hays, Kansas. Through videoconferencing, neuropsychologists in

Kansas City were able to interview the patient and their family/caregivers. The neuropsychologist then conferred with the remote psychometrician and observed test administration through videoconferencing. Upon interpretation of results by the neuropsychologist, feedback was provided through videoconferencing with the report being faxed to the referral source (Tröster et al., 1995). Other comparable studies utilized videoconferencing to complete intake evaluations and provide feedback, as opposed to direct administration of measures. Similar to this study, Ball & Puffett (1998) utilized a remote assistant to supply participants with stimuli needed to complete tasks as part of a structured interview (i.e., the CAMCOG) to determine if individuals had early cognitive impairment. The structured interview was conducted over videoconferencing and the assistant was in a distant location with the patient. Clement et al., (2001) utilized videoconferencing to provide neuropsychological evaluation feedback to remote military personnel. Remote evaluations of cognitive functioning found 100 percent consensus amongst clinicians when compared to face-to-face interviews (Shores et al., 2004). The use of videoconferencing for clinical interviews and feedback has occurred semi-regularly. As it was determined videoconferencing was a sufficient format for information gathering and allowed for accurate clinical impressions, research was extended to look at completing the entire evaluation, including testing, to determine cognitive functioning of patients.

The first study typically cited in meta-analytic and systemic reviews to use videoconferencing to administer assessments for neuropsychological purposes was conducted in 1997 (Montani et al.). This study assessed fifteen geriatric patients (*M* = 88 years old) using the Mini-Mental Status Exam and Clock Face Drawing to determine cognitive functioning. An interview was conducted over video conferencing, with the testing following the interview. This initial experiment showed there was a significant difference in performance when comparing face-to-face and video-conferencing administration (Montani et al., 1997). Although, patient satisfaction may have been the decisive factor to such an extent that it overshadowed clinical implementation. The majority of participants felt telecommunication was impersonal and they reported experiencing hearing difficulties related to the volume of audio transmission (Montani et al., 1997). With that said, most early and subsequent studies notated that participants were not deterred by the use of technology and felt the interactions were appropriate and efficient (Cullum et al., 2006; Hildebrand et al., 2004; Kirkland, Peck, & Bennie, 2000; Menon et al., 2001; Vestal et al., 2006). In studies assessing length of appointments, some studies found a significant difference between tests with some notating videoconferencing took longer (Cullum et al., 2014; Kirkland et al., 2000) and alternatively some finding teleneuropsychology testing required shorter appointments or were not-significantly different when assessed for time (Parikh et al., 2013).

Additional teleneuropsychology studies expanded the breadth of cognitive testing administered through videoconferencing, studied the difference in consultation length, and evaluated participants' impressions and preferences of test administration modality. Tests that have been utilized and assessed include the National Adult Reading Test (NART), Matrix Reasoning subtest, Quick IQ Test, Adult Memory and Information Processing Battery (AMIPB), Benton Visual Retention Test, Hopkins Verbal Learning Test-Revised (HVLT-R), Logical Memory, Oral Trails, Brief Test of Attention (BTA), Digit Span, Seashore Rhythm Test, Activities of Daily Living (ADL), Instrumental Activities of Daily Living (IADL), Mini Mental State Examination (MMSE), Boston Naming Test (BNT), Controlled Word Association Test (CWAT), Multilingual Aphasia Examination (MAE), Phonemic fluency, Picture Descriptions, Semantic fluency, Vocabulary subtest, Clock Drawing (CD), Silhouette subtest, Symbol Digit Motor Test, Grooved Pegboard, Geriatric Depression Scale (GDS), and Informant Questionnaire for Cognitive Decline in the Elderly (IQCODE) (Carotenuto et al., 2018; Ciemins et al., 2009; Cullum et al., 2006; Cullum et al., 2014; Grosch et al., 2015; Hildebrand et al., 2004; Jacobson et al., 2003; Kirkland et al., 2000; Lindauer et al., 2017; Loh et al., 2004; Loh et al., 2007; Vestal et al., 2006; Wadsworth et al., 2016; Wadsworth et al., 2018).   Results from these studies are available for review in Table 1 below. Testing has since expanded with more commonly used assessments being included in teleneuropsychological evaluations.

Table 1. Comparison of tele- and in-person neuropsychological assessment results.

Study	**Measure**	***r***	**ICC *r***	***p*-value**	**Limits of Agreement**
Abodolahi et al., 2016 (*n* = 17)	MoCA	.59	.59		
Carotenuto et al., 2018 (*n* = 28)	MMSE			.37-.68	
	ADAS-cog			.07-.19	
Chapman et al., 2019 (*n* = 24)	MoCA		.62	.658	-5.2 to +4.9
Ciemins et al., 2009 (*n* = 73)	MMSE	≥.86			
Cullum et al., 2006 (*n* = 33)	MMSE	.89	.88		
	HVLT-R	.55-.77	.54-.77		
	Digit Span	.81	.78		
	Category Fluency	.58	.58		
	Letter Fluency	.83	.83		
	BNT	.88	.87		
Cullum et al., 2014 (*n* = 202)	MSSE		.91	.93-.99	
	HVLT-R		.80**	.01-.79	
	Letter Fluency		.85	.34-.79	
	Category Fluency		.72	.05-.38	
	BNT-15		.81**	.003-.073	
	Digit Span Forward		.59	.17-.30	
	Digit Span Backward		.55	.60-.84	
	Clock Drawing Test		.71*	.04-.74	
Galusha-Glasscock et al., 2015 (*n* = 18)	RBANS		.59-.90		
Grosch et al., 2015 (*n* = 8)	MMSE		.42		
	Clock Drawing Test		.44		
	Digit Span		.72		
Harrell et al., 2014*** (*n* = 31)	MoCA				
	WTAR				
	Digit Span				
	Trail Making Test				

Table 1 (continued).

Study	Measure	*r*	ICC *r*	*p*-value	Limits of Agreement
Harrell et al., 2014 continued	BNT				
	COWA				
	Animal Naming				
	RCFT				
	WMS LM				
	HVLT-R				
	BVMT-R				
	D-KEFS Proverbs				
	PHQ-9/GDS/GAD-7				
	TOMM				
	MMPI-2-RF				
	PTSD Checklist				
	WRAT-4				
	CVLT-II				
	BDAE				
	TOP-J				
	ILS				
Hildebrand et al., 2004 (*n* = 29)	RAVLT				-5.70 to +6.72
	BTA				-5.09 to +6.95
	Matrix Reasoning				-4.56 to +6.08
	Vocabulary				-3.07 to +3.13
	CWAT				-4.34 to +4.82
	Clock Drawing Test				-22.07 to +18.21
Jacobson et al., 2003 (*n* = 32)	Pegboard	.71-.83		.19-.35	
	VOSP Silhouettes	.64		.84	
	BVRT	-.62-.64		.27-.63	
	SDMT	.37-.69		.77-.83	
	Seashore Rhythm Test	.77*		.03	
	Digit Span	.82		.33	
	Vocabulary	.86		.83	
	WMS LM	.79-.82*		.02-.17	
Kirkland et al., 2000 (*n* = 27)	NART				-4.57 to +4.87
	Quick IQ Test				-19.96 to +12.92
	AMIPB				-33.93 to +37.55
Lindauer et al., 2017 (*n* = 28)	MoCA		.86-.93		
	CDR		.75		
	RMBPC		.77		
	GDS		.67		
Loh et al., 2004 (*n* = 20)	MMSE	.90			
	GDS	.78			
Loh et al., 2007 (*n* = 20)	MMSE		.89		
	GDS		.78		
	IQCODE		.88		
	ADL		.96		
	IADL		.88		
Montani et al., 1997 (*n* = 15)	MMSE	.95**		.003	
	Clock Drawing Test	.55		.07	
Vahia et al., 2015 (*n* = 22)***	MMSE (*Spanish)*				
	HVLT-R (*Spanish*)				
	Letter Fluency (*Spanish*)				

Table 1 (continued).

Study	Measure	*r*	ICC *r*	*p*-value	Limits of Agreement
Vahia et al., 2015 continued	Category Fluency (*Spanish*)				
	Clock Drawing Test (*Spanish*)				
	BVMT-R (*Spanish*)				
	Digit Span (*Spanish*)				
	Ponton-Satz Spanish Naming Test				
Vestal et al., 2006 (*n* = 15)	BNT			.864	
	Token Test			.279	
	Picture Descriptions			1.000	
	Aural Comprehension of Words and Phrases			.230	
	COWA			.188	
Wadsworth et al., 2016 (*n* = 84)	MMSE		.92	.133	
	Clock Drawing Test		.65	.096	
	Digits Forward		.72**	.004	
	Digits Backward		.69	.760	
	HVLT-R		.84-.90**	.005-.038	
	Letter Fluency		.93	.920	
	Category Fluency		.74*	.026	
	Oral Trails		.79-.83**	<.001-.726	
	BNT		.93**	.002	
Wadsworth et al., 2018 (*n* = 197)	Clock Drawing Test			.520	
	Digit Span Forward			.276	
	Digit Span Backward			.635	
	BNT-15			.806	
	HVLT-R			.457-.735	
	Letter Fluency			.814	
	Category Fluency			<.001**	

Note: * $p < .05$, ** $p < .01$,

Note: ***Harrell et al., looked at patient acceptance of assessment over videoconferencing and only notated that "standardized administration instructions are provided for all measures, with slight modifications on the Trail Making Test" (2014, p. 3). Vahia et al., indicated there were no significant differences on "individual cognitive tests" (2015, p. 669).

Acronyms: ADAS-cog = Alzheimer's Disease Assessment Scale cognitive subscale; ADL = Activities of Daily Living; AMIPB = Adult Memory and Information Processing Battery, story recall, list learning, figure recall, and information processing; BDAE = Boston Diagnostic Aphasia Exam; BNT = Boston Naming Test; BVMT-R = Brief Visuospatial Memory Test – Revised; BVRT = Brief Visual Retention Test; BTA = Brief Test of Attention; CDR = Clinical Dementia Rating; COWA/CWAT = Controlled Oral Word Testing Association Test (FAS); CVLT-II = California Verbal Learning Test – II; D-KEFS Proverbs subtest = Delis-Kaplan Executive Function System, Proverbs subtest; GDS = Geriatric Depression Scale; HVLT-R = Hopkins Verbal Learning Test-Revised; ILS = Independent Living Scales, Health and Safety subtests; IADL = Instrumental ADL assessment; IQCODE = Informant Questionnaire for Cognitive Decline in the Elderly; MMPI-2-RF = Minnesota Multiphasic Personality Inventory – 2-RF; MMSE = Mini-Mental Status Exam; MoCA = Montreal Cognitive Assessment; NART = National Adult Reading Test; PHQ-9 = Patient Health Questionnaire – 9; RAVLT = Rey Auditory Verbal Learning Test; RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; RCFT = Rey-Osterreith Complex Figure; RMBPC = Revised Memory and Behavioral Problems Checklist; SDMT = Symbol Digit Motor Test; TOMM = Test of Memory Malingering; TOP-J = Test of Practical Judgment; VOSP Silhouette = Visual Object and Space Perception Battery, Silhouette subtest; WMS-IV LM = Wechsler Memory System-IV, Logical Memory I and II; WTAR = Wechsler Test of Adult Reading; WRAT-4 = Wide Range Achievement Test – 4, Reading and Spelling subtests

In the last ten years, studies have expanded their assessment regimens to administer larger batteries over videoconferencing. Results also further bolstered the support of the aforementioned assessments for administrating cognitive measures over videoconferencing. Additional measures assessed included the Brief Visuospatial Memory Test – Revised (BVMT-R; Vahia et al., 2015), Clinical Dementia Rating Scale (CDR; Lindauer et al., 2017), Memory and Behavioral Problems Checklist (RMBPC; Lindauer et al., 2017), Montreal Cognitive Assessment (MoCA; Abodolahi et al., 2016; Chapman et al., 2019; Lindauer et al., 2017), and Repeatable Battery of the Assessment of Neuropsychological Status (RBANS; Galusha-Glasscock et al., 2015). Harrell et al., (2014) administered a variety of assessments within a remote neuropsychological clinic. These included California Verbal Learning Test (CVLT-II), Wechsler Test of Adult Reading (WTAR), Rey-Osterreith Complex Figure, Proverbs Subtest, Test of Memory Malingering (TOMM), Wide Range Achievement Test (WRAT-4), Independent Living Scales (ILS), and Test of Practical Judgment (TOPJ). Despite this, Harrell et al., did not report any data supporting the use of these measure through videoconferencing and only noted that families that participated in the services provided "highly positive" general and unsolicited feedback (2014). The premise of their research looked at the additive value of neuropsychological evaluations rather than judging the validity and reliability of the assessment measures (Harrell et al., 2014). With the recent COVID-19 pandemic, the focus on teleneuropsychological research has increased significantly as neuropsychologists attempted to maintain their practices while prioritizing patient safety and health.

Overarchingly, studies have found non-significant differences between face-to-face and videoconferencing administration (Cullum et al., 2014; Hildebrand et al., 2004; Loh et al., 2007; Kirkland et al., 2000; Vestal et al., 2006; Galusha-Glasscock et al., 2015; Wadsworth et al., 2018). Those from the above list that occasionally produced significantly different performances when comparing face-to-face testing and videoconferencing include Adult Memory and Information Processing Battery (AMIPB; Kirkwood et al., 2000), Ammon's Quick Test (Kirkwood et al., 2000), Boston Naming Test (BNT; Wadsworth et al., 2016), Clocking Drawing (CD; Cullum et al., 2006; Galusha-Glasscock et al., 2015; Hildebrand et al., 2004), Digits Forward (Wadsworth et al., 2016), Hopkins Verbal Learning Test – Revised (HVLT-R; Cullum et al., 2006; Cullum et al., 2014), Logical Memory I (from Weschler Memory Scales - Revised; Jacobson et al., 2003), Oral Trail Making – A (Wadsworth et al., 2016), Semantic Fluency (Wadsworth et al., 2018), and Seashore Rhythm Test (Jacobson et al., 2003). Despite these results, other studies found most of these measures' performance across administration modality to be adequate (Brearly et al., 2017; Cullum et al., 2006; Cullum et al., 2014; Wadsworth et al., 2016; Wadsworth et al., 2018). In a meta-analysis, Boston Naming Test was the only assessment that maintained a small but significant difference in administration modalities, with videoconferencing producing a lower score (Hedges g = -0.12; $p < 0.05$; Brearly et al., 2017; Marra et al., 2020a). These findings suggest that Boston Naming Test may produce lower scores in participants assessed through videoconferencing and that must be considered when interpreting these results. Many studies have been conducted evaluating the usefulness and appropriateness of administering cognitive measures over videoconferencing, but neuropsychological evaluations also consist of assessments outside of traditional measures of cognition.

Two such areas lacking empirical evaluation are performance validity tests and social cognition measures. To our knowledge, the validity of these measures has never been assessed through videoconferencing within the context of a neuropsychological evaluation. Performance validity tests (PVT) are empirical measures that assist neuropsychologists in determining whether a patient's performance is valid (i.e., the results are indicative of the patient's true cognitive functioning) or invalid (i.e., the results were impacted by external/internal motivations, patient's effort, or disengagement in the assessment) (Greher & Wodushek, 2017a; Greher & Wodushek, 2017b). These measures can either be standalone measures that were developed strictly to measure validity (e.g., Test of Memory Malingering; TOMM) or embedded within another measure (e.g., Forced Choice within the California Verbal Learning Test; Martin, Schroeder, & Odland, 2015). Additionally, social cognition has not been evaluated within a neuropsychological evaluation through videoconferencing. Social cognition addresses numerous areas of social and cognitive psychology, such as impressions and attitudes that inform an individual's decisions and judgments (Hunt, Borgida, & Lavine, 2012). These ideas are often referenced within the terms of attributional bias, emotion processing, social perception, and theory of mind (Javed & Charles, 2018). Not only are these skills negatively impacted within a variety of psychiatric and personality disorders, an individual's functioning can initially be poor or deteriorate within the context of a neurodevelopmental (i.e., Autism Spectrum Disorder) or neurodegenerative disorder (i.e., Frontotemporal Disease; Baron-Cohen et al., 2015; Hutchings et al., 2015). The purpose of this study is to assess the appropriateness of PVT and social cognition measures administered through videoconferencing.

Teleneuropsychology Guidelines

In 2011, Grosch et al., produced a series of guidelines regarding the emerging field of teleneuropsychology. These guidelines detailed the importance of informed consent, privacy, confidentiality, competence, licensure, reimbursement, technological devices, and availability of staff during testing. Clinicians are expected to inform their patients of any modified services;

this is especially important when discussing the limits of privacy and confidentiality as information is transmitted over the internet. Given the use of technology, patients should be made aware that additional technicians may be necessary to assist with troubleshooting if issues occur with the videoconferencing sessions. Reiterating the American Psychological Association's Ethics Code (2002), it is important that clinicians are competent prior to using videoconferencing to ensure they are protecting others from any undue harm. This plays an especially important part when neuropsychologists are determining what assessments to utilize as one must consider the modifications and alterations that may influence results when administered through a videoconferencing platform. Addressing assessment limitations is very important in the final evaluation, as well as documenting any interruptions that may occur through videoconferencing. Grosch et al., (2011) detailed the importance of using high-quality videoconferencing equipment allowing for an interactive encounter. It was also suggested that a staff member be available at the remote location to provide supplies and assist the patient with any difficulties that arise. As the practice of using videoconferencing within other facets of psychology increased, professional organizations produced guidelines that mirrored the above suggestions but expanded the discussion to include all provider-client interactions.

The American Psychological Association dispersed guidelines related to telepsychology and its utilization in clinical practices in 2013. While the breadth of these guidelines touched on all facets of psychology, the portions dedicated towards assessment detailed the importance of maintaining the integrity of the instruments, limitations of test administration, and suitability of assessments given through videoconferencing (American Psychological Association, 2013). The American Psychiatric Association and American Telemedicine Association (ATA) published their best practices regarding videoconferencing-based mental health in 2018. The Association of State and Provincial Psychology Boards (ASPPB) detailed each state's response to COVID-19 and available opportunities for temporary licensures which would allow for the continued use of teleneuropsychology (ASPPB, 2020). In response to COVID-19, the Inter Organizational Practice Committee (IOPC; Bilder et al., 2020) published guidelines geared towards promoting and continuing neuropsychological care in the midst of the pandemic. The IOPC (Bilder et al., 2020) addressed important aspects of the informed consent, privacy, determination of patient appropriateness for videoconferencing, test modification, Health Insurance Portability and Accountability Act (HIPPA) compliant technology requirements, environmental concerns, and potential for lost richness of observations. Although this research project emphasizes teleassessment, it is imperative to recognize that assessment is just a part of a neuropsychological evaluation; diagnostic interviews, review of medical records, and feedback all are portions of the evaluation that can be readily performed through videoconferencing or available technologies (Cullum et al., 2019; Stolwyk & Hammers, 2020).

A variety of concerns are often discussed when determining if videoconferencing is an appropriate format for administering assessment. These included environmental distractors, test modification, and assessment integrity. The American Psychological Association (2013) urged clinicians to be conscious of environmental conditions that could impact the patient's testing (e.g., access to the internet, being coached by someone off camera, or various other distractors that exist in someone's home). The IOPC discussed the importance of placing the patient in a quiet area of the home without family members or pets (Bilder et al., 2020). Throughout the appointment, the provider should document any disruptions or technological difficulties (American Telemedicine Association, 2018; Grosch et al., 2011; Hewitt et al., 2020). Bilder et al., (2020) suggested the use of headphones to assist in eliminating distractions and that patients should hide their view of themselves on the screen so that they are not distracted by their image. The American Psychological Association noted the importance of expressing which measures were provided over telepsychology and to dictate any modifications or accommodations that were made so that those reading the evaluation are aware. When presenting test stimuli, it is important that clinicians utilize the “Share Screen” feature as a higher quality and more consistent image is presented to the patient (Bilder et al., 2020). If a patient is asked to complete visual memory tasks, it is imperative the provider observe the patient placing their produced image in a folder or out of their vantage point so as to maintain as close to standardized testing procedures as possible (Bilder et al., 2020; Grosch et al., 2011). When holding a teleassessment appointment, it is essential that clinicians disable recording abilities on the telehealth platform to protect test security. Combining all of the above guidelines, three models of teleneuropsychology clinics have emerged. It should be noted that an evaluation of models implanted in response to COVID-19 has yet to be released.

Models of Teleneuropsychology

When discussing how one would implement teleneuropsychology into their own practice, clinicians must determine what model fits best with their clientele and their chosen test battery. Prior to COVID-19, most remote centers that considered themselves teleneuropsychology clinics utilized a trained psychometrician who was present with the client during testing; testing was also observed by a remote neuropsychologist (Ball & Puffett, 1998; Tröster et al., 1995). Interviews are conducted through videoconferencing with the neuropsychologist, but testing is completed with a trained psychometrist who is supervised by the neuropsychologist (Clement et al., 2001; Kane & Parsons, 2017). This allowed materials to be manipulated by the patient and in-person behavioral observations to be provided to the clinician. This also maintains the standardization of testing and allows for a wider breadth of measures to assist with clinical diagnoses; this seems to be a logical extension of neuropsychology's current practices but does not fully envelope the freedoms of remote assessment. Utilizing a "trained assistant" was encouraged to help support the patients at the off-site location, but also to better the security of the assessment process and materials (American Psychological Association, 2013). The model described above is utilized in some areas, including the University of Utah's Teleneuropsychology Program (Stolwyk & Hammers, 2020). Benefits of this particular model include comparable batteries and allow for a wide range of severity within patients seen in the clinic. With this model, patients are still expected to travel long distances to complete testing and staffing trained technicians can be difficult (Stolwyk & Hammers, 2020).

Another model that has been proposed is the utilization of numerous remote sites that still require patients to travel, but distance is often shortened. These remote sites require the assistance of a psychometrician or similar assistant, but testing administration is completed through videoconferencing (Cullum et al., 2019; Stolwyk & Hammers, 2020). Ultimately, this allows less trained individuals to be present with the patient as they may assist with technological difficulties and they can provide the patient with forms, test stimuli, or pencils (Kane & Parsons, 2017). Given the more hands-off approach, patients must be more cognitively intact and the battery may be limited, as tangible tasks are not used (Cullum et al., 2020). Additionally, this model also raises questions about ethical constraints as related to licensure. For example, if a remote site is located outside of the clinician's practicing state, the patient would not be completing the assessment in the same jurisdiction. This is a problem because there is currently no recognized national licensure organization in which all fifty states have enforced legislation for its use. It would be unethical, as legislation stands currently, for the clinician to see patients residing outside of their state of licensure while conducting the evaluation. This issue remains when the third model is discussed where patients remain in their home during the evaluation.

The final model, which is the one utilized in this study, is a completely remote process. Participants are in their own homes, using their own equipment, during the appointment. Provided this is appropriate for the patient, the individual must be relatively high functioning as they will need to be able to troubleshoot as needed (Kane & Parsons, 2017; Stolwyk & Hammers, 2020). Given current literature constraints, the assessment battery that is used for this evaluation is significantly restricted as few assessments have been determined to be reliable across administration modalities (Brearly et al., 2017; Marra et al., 2020a). With that said, many neuropsychological assessments are orally administered and only require an oral response, making the transition to telehealth relatively easy given the allowance for near-real-time interactions (Kane & Parsons, 2017). Additionally, this raises questions concerning the validity of the testing as there is little guarantee that the patient is abiding by standard testing procedures (Cullum et al., 2020; Stolwyk & Hammers, 2020). The guidelines established by the IOPC attempt to address this through utilizing as many public domain tests as possible, not mailing materials to patients, and monitoring the patient's environment. In three studies assessing the validity and effectiveness of teleneuropsychological evaluations conducted at home (Abdolahi et al., 2016; Lindauer et al., 2017; Stillerova et al., 2016), this was not observed to be an issue; regardless, providers should be aware of these possible concerns and mitigate them to the best of their ability.

Regardless of the model selected, when establishing a teleneuropsychology practice, the practitioner must have certain technological features (e.g., computer with camera and necessary bandwidth / internet capabilities). Through the use of screen sharing or dedicated cameras, the examiner should be able to provide the participant with visual stimuli as needed. Some studies have had material stationed in front of a camera that was then viewed by the participant (Galusha-Glasscock et al., 2016; Wadsworth et al., 2016), whereas more advanced technological innovations allow materials to be scanned into PDF form and shared on the computer screen (Bilder et al., 2020). It is important throughout the evaluation that the examiner be able to view the participant so as to monitor attitude, motor behaviors, and other relevant behavioral observations. Considering the use of technology, all models of teleneuropsychology must determine that the technologies used are HIPAA compliant, just as they would be if the patient was seen in their office. Additionally, there must be a telehealth consent form provided to the patient as this telehealth process holds additional concerns that patients must acknowledge (Bilder et al., 2020; Grosch et al., 2011; Hewitt et al., 2020). Clinicians must also consider how they would handle an emergency if it arose so it is important that the location of the patient is established during the initial consent section. Many of these considerations and models were put into action with the onset of COVID-19 when numerous stay-at-homes orders were established. Neuropsychologists were placed in a pressing predicament where they were expected to rapidly meet the needs of their patients under the constraints of teleservices.

Neuropsychologists' Response to COVID and Recent Practices

An international survey was collected during a continuing education presentation in April 2020, shortly after the onset of COVID-19 by Hammers and colleagues (2020). This survey evaluated the provision and anticipated use of teleneuropsychological services in response to COVID-19. Results indicated 11-28% of practitioners used videoconferencing to provide services to their patients prior to the COVID-19 pandemic; of those individuals who endorsed this practice, only 11 percent of clinicians utilized teleassessment, most of the services provided through this modality were the clinical interview or feedback. As of April 2020, when this survey was conducted, these practices had increased to 15-52 percent of providers offering clinical interviews, testing, feedback, or interventions through videoconferencing (Hammers et al., 2020). In part due to the teleneuropsychological research consolidated by Brearly et al. (2017) and Marra et al. (2020a), between 59-90 percent of these same providers anticipated their use of videoconferencing for various portions of their evaluation will increase (Hammers et al., 2020). A subsequent survey conducted by Marra et al., (2020b) reported similar results as 14-51% of clinicians reported using telehealth services currently for their neuropsychology practices. Despite this increased necessity for teleneuropsychology, there is apprehension related to the validity and feasibility of this practice long term.

As identified by the IOPC committee, there are appropriate concerns related to competency, licensure, and reimbursement surrounding teleneuropsychology. Numerous webinars have been published over the last six months (at the time of this writing), that discuss standardization, technical information/equipment, and risk management. It was recommended by the IOPC that providers utilizing teleneuropsychology participate and watch so as to inform their own practices (Bilder et al., 2020). Prior to COVID-19, each state required clinicians to be licensed within the state in which they practiced and in whatever state patients resided if telehealth sessions were occurring. In the midst of COVID-19, these stipulations were relaxed for some states, allowing temporary licensing to occur across state lines (ASPPB, 2020). This has led to further tailoring of the Psychology Interjurisdictional Compact (PSYPACT) which is a reciprocal compact that allows clinicians to practice in a handful of states so long as they are licensed in one (ASPPB. 2020). Currently 34 states have enacted PSYPACT legislation, with 1 state having enacted PSYPACT but the legislation has not yet gone into effect, and an additional 2 states have pending PSYPACT legislation (PSYPACT, 2022).

Prior to COVID-19, neuropsychological evaluations conducted through a videoconferencing format did not have appropriate billing codes to allow for reimbursement; it was the responsibility of the clinician to contact insurance companies to determine if the services would be covered (Bilder et al., 2020; Grosch et al., 2011). Historically, the clinical interview held prior to test administration was covered, but assessments administered over videoconferencing were not covered (Bilder et al., 2020). In support of the necessity of telehealth services, the U.S. Congress passed the Coronavirus Aid, Relief, and Economic Security (CARES) Act in March 2020, which temporarily authorized neuropsychology CPT® codes to be reimbursed through Medicaid and many private insurers (American Psychological Association, 2020; Bilder et al., 2020; Marra et al, 2020b). The American Psychological Association is working with insurance companies to ensure reimbursement for teleassessment is continued subsequent to the COVID-19 pandemic (American Psychological Association, 2020; Hammers et al., 2020; Marra et al., 2020b).

Two reviews have been conducted covering the field of teleneuropsychology; studies were only included if they did not utilize a trained technician as testing was completed entirely through videoconferencing, similar to the current study. The first was a meta-analysis completed in 2017 (Brearly et al.), and the most recent was conducted in response to the COVID-19 pandemic in an effort to help neuropsychologists find their footing in a world of technological advances that they had not previously embraced (Marra et al., 2020a). When looking at all measures assessed, Brearly et al., (2017) found that there was not a significant change in test scores based on administration modality with only a small effect (Hedge's $g = -0.03$; $p > .05$).

Secondary analyses found that videoconferencing administration of non-synchronous dependent tests (those that did not require timing or limited repetition) produced scores approximately 1/10th of a standard deviation lower when compared to face-to-face administration (Hedge's $g = -0.10$); this was determined to be a significant but small difference ($p < .001$; Brearly et al., 2017). On synchronous-dependent measures that were timed or utilized repetition, there were no significant differences between administration modalities (Brearly et al., 2017). Seemingly in response to COVID-19, Marra et al., (2020a) published an article furthering the research of Brearly et al. (2017) but included additional studies to inform clinicians as they created their own tele-practices during the pandemic. The stance and results from both reviews suggest videoconferencing is a feasible method of assessment administration when limited options for face-to-face testing are available.

Despite the growing literature on teleneuropsychology, there are no standardized norms established for providers to use when assessments take place via videoconferencing. The idea of completing assessment with some modification to standardization is not foreign to neuropsychologists. One could compare this to completing in-hospital or bedside evaluations, which is within the scope of a neuropsychologist's job. Within these parameters, conservative interpretation and determination of appropriate norms is imperative despite the promising research that there is little difference between face-to-face and videoconferencing administration (Cullum et al., 2019; Cullum et al., 2020). Almost all of the studies assessing the feasibility of teleneuropsychology have been conducted in controlled clinic settings, with limited studies occurring in the patients' homes (Cullum et al., 2019; Cullum et al., 2020; Stolwyk & Hammers, 2020). Observing the above information, it is increasingly important that the field of neuropsychology further their research in determining the appropriateness of administering cognitive assessments through videoconferencing. As previous research, experts, and meta-analyses show, there are still large gaps in the literature and the present COVID-19 pandemic has had made these needs glaringly apparent.

Summary of Study Purpose and Related Hypotheses

The purpose of this current project is to determine the feasibility of performance validity testing and social cognition measures through the use of videoconferencing. The results of this study could provide the field of neuropsychology the opportunity to expand the use of current assessments and offer its services to those without readily available neuropsychological services through telehealth. With the COVID-19 pandemic, clinicians around the world have raced to piece together batteries that would support their clinical diagnoses while maintaining standardization through videoconferencing; the literature concerning teleneuropsychological assessment is still relative bare at this point when compared to telehealth utilization in other healthcare fields. By developing teleneuropsychological services, this allows the field of neuropsychology to further enter the digital age, maximizing the available technologies to best serve the public. Although there are numerous benefits to further understanding the benefits of teleneuropsychological practices, one of the greatest developments from this pandemic has been the insurances' inclusion of coverage for teleassessment. This allows families with individuals suffering from a variety of neurodegenerative and neurodevelopmental symptoms to not forego services while the government determined what services were essential or unessential. The purpose of this study not only builds upon current literature but furthers the teleneuropsychology practice by two ways. The first is that this study is being conducted with participants placed in their home. Secondly, this study expands the literature on cognitive, performance validity, and social cognitive measures that have been studied previously when administered through videoconferencing.

Individuals included in this study were recruited through the use of the SONA System and seen at a Midwestern University; this group will then be compared to a control group from the same university a study where data were previously collected. A select core of cognitive, performance validity, and social cognition measures will be administered to participants. Given the constraints of COVID-19 and to ensure the safety of both examiners and participants, scores will be compared to comparison groups that were administered the same measures through face-to-face administration at the same Midwestern University. Six different hypotheses will be addressed by this study.

1. There will be no significant differences observed between face-to-face testing and videoconferencing on a variety of cognitive instruments. Specifically, performance on the North American Adult Reading Test, Oral Trails B, California Verbal Learning Test – II, Verbal Fluency, Digit Span, and Raven's Progressive Matrices, will not show a significant difference in performances between individuals who were administered the assessment through videoconferencing when compared to testing norms of individuals in a face-to-face setting. Not only will across group comparisons show no deficits in cognitive testing, but no significant differences will be noted in individuals' performances on stand-alone and embedded performance validity testing.
2. No significant differences will be found between in person and telehealth performance on the performance validity tests. Specifically, individuals administered the Test of Memory Malingering, Dot Counting Test, Forced Choice on CVLT-II, and Reliable Digit Span

through videoconferencing will perform comparable to testing norms of individuals in a face-to-face setting.

3. No significant difference will be determined when a social cognition measure is administered face-to-face or through videoconferencing. Of note, Reading the Mind in the Eyes will produce no significant differences across groups. Individuals who are administered the social cognition measures through videoconferencing will perform at a non-significantly lower level than testing norms established through a face-to-face setting.
4. A positive correlation will be found between the National American Adult Reading Test and Raven's Progressive Matrices in assessing overall intellectual functioning. For this study will we be using an abbreviated version of the Raven's with little data supporting its validity.
5. A significant difference will be observed between Oral Trails A and Trail Making Test A. Individuals who perform Oral Trails will take significantly less time to complete the task than those in the control group. These results are suggested by the literature (Axelrod & Lamberty, 2006) and this will be mirrored within this study.
6. Matrix reasoning and vocabulary will both be significant predictors of cognitive empathy, an aspect of social cognition, even after controlling for relevant covariates.

CHAPTER THREE

METHODOLOGY

Study Participants

Participants were recruited from a Midwestern University (*N* = 532) through the use of SONA experiment management software. 401 individuals were previously collected with 131 participants collected for this specific research question. This number is based off a simulation assessing study sample size recommendations for multiple hypothesis testing (Bujang, Sa'at, Joys, & Ali, 2015). Individuals received compensation in the form of credit for undergraduate or graduate courses. Inclusion criteria for all six hypotheses required participants to have access to a computer, laptop, tablet, or other device with videoconferencing capabilities (e.g., video camera and microphone). Additionally, participants were required have access to the internet and be between the ages of 18 and 89 in order to participate in the study. Two individuals were excluded from data analyses due to not completing the study and three additional individuals were excluded on particular analyses due to not completing certain tasks in their entirety (e.g., F-Fluency, Dot Counting Test, Digit Span, Raven's Progressive Matrices, and Oral Trails B). The final total number of participants included in analyses from the teleadministration group was 129, with some analyses only including 128 participants to account for the individuals that did not complete certain tests.

The teleneuropsychology sample (Group 1) was predominantly White (62.0 percent) and female (63.6 percent). Eight individuals identified themselves as transgender or nonbinary. For the purpose of this study, only one measure utilized gendered norms (CVLT-II). As no significant differences were found when comparing their performances when scored using both male and female norms, these individuals' scores were coded with female norms as the majority of our sample self-identified as female (Webb, Holmes, & Peta, 2016). The average age of the sample was 21.58 years old (*SD* = 4.75; with a range of 18 to 41) with 14.36 years of education (*SD* = 1.80). The majority of individuals utilized laptops (80.6 percent) although desktop computers (13.2 percent), tablets (2.3 percent), and other devices that had video and audio capabilities (3.9 percent) were used as well. Of individuals in the full teleadministration group (*N* = 131), 15.3 percent experienced technical difficulties (i.e., audio problems, poor internet connection, or complete disconnection), with 2.3 percent of individuals (*N* = 3) unable to complete the study in its entirety due to these issues. Of these three individuals, two lost connection after the first task (approximately five to ten minutes into the study) and they were excluded from analyses. The other individual was able to complete the majority of the study with the exception of the final three measures; this person's data were only included in the analyses of the tasks that they completed.

The first comparison group (Group 2) contained a population of students from the same Midwestern University and the study was previously approved by the University's Institutional Review Board. The first comparison sample collected a sample 100 of participants, although only the 50 individuals identified as the control group were used for this study's analysis. Of this sample, they were primarily Caucasian (62.0 percent) and female (72.0 percent). The average participant age was 19.92 years old (*SD* = 1.90) with 14.10 years of education (*SD* = 1.44). Inclusion criteria required participants to be a university student between the ages of 18 and 24 years old. Participants were recruited through SONA experiment management software and were compensated with course credit for participating in research.

Reading the Mind in the Eyes required the use of another comparison group (Group 3); those data were also collected from the same Midwestern University and approved by the

University's Institutional Review Board. The participants (N = 181) ranged in age from 18 to 52, with the average age of 21.58 years (*SD* = 5.49). Once again, most participants were primarily White (56.9 percent) and female (62.2 percent). The average level of education was 13.51 years (*SD* = 3.9 years). Participants were recruited through SONA systems and may have received course credit or extra credit for participating in research. Participants needed to be a student at the university, able to give consent, and be between the ages of 18 and 89 years old.

Raven's Progressive Matrices Shortened Form again required the use of another comparison group (Group 4); those data were also collected from the same Midwestern University and approved by the University's Institutional Review Board. As compared to previous comparison groups, this study's sample was comprised of 171 students who were primarily female (68.4 percent) and White (56.7 percent). It was required that the participants be Native American English speakers, be between the age of 18 and 89 years old, and attend the university. Participants were recruited through SONA systems and may have been compensated for their involvement through course credit or extra credit.

Procedures

Data collection for the teleneuropsychology sample was completed in its entirety through Zoom Communications videoconferencing technologies and utilized PowerPoint for administration of assessment stimuli when necessary. The SONA advertisement for this study specified the requirement that potential participants have access to a computer with a video camera and microphone capabilities. Upon scheduling a time to meet with the examiner over videoconferencing, participants were consented into the study. Participants were instructed to locate themselves in a space with minimal distractions so as to maximize their engagement with the experiment. Audio levels were adjusted accordingly so that conversation-level speaking could be appropriately heard.

The examiner presented necessary assessment materials in a fixed position through the screen share feature on Zoom. PowerPoints were created prior to administration which allowed the examiner to switch the stimuli by changing the visible slide, which was intended to mimic the process of turning pages while administering face-to-face batteries. The lead examiner has used this technique in clinical practice and has encountered no problems in administration. The tests that required the utilization of PowerPoint included the National American Adult Reading Test, Test of Memory Malingering, Dot Counting Test, and Reading the Mind in the Eyes. Administration of the assessments occurred in the same order for all participants to limit any confounds related to test order and to mirror that of the comparison groups (with minor alterations to account for delays). The order was as follows: The National American Adult Reading Test, California Verbal Learning Test, Test of Memory Malingering, Digit Span, Oral Trails, Reading the Mind in the Eyes, Raven's Progressive Matrices, Dot Counting Test, and Verbal Fluency.

Although counterbalancing of assessments occurs in other studies, this is not frequently conducted in real world clinical practices as test delays often negate the functionality of counterbalancing (Rabin et al., 2016; Sweet et al., 2011). For example, the CVLT-II used in this assessment requires two delays, one for 20-minutes and another for 10-minutes. It is commonly practiced that non-interfering assessments are administered during that time which will differ based on the referral question as well as well as their performance on earlier exams. Given the amount of testing, the study took approximately one-and-one-half hours for the participants to complete all testing. Participants did not receive the results of their testing and they were allowed to ask questions at the end of the study.

Participants in Group 2 were consented into the study and randomly designated to one of two groups; they were instructed to give their best effort or were coached to malinger. As we only want true performance comparisons, only those from the control group who were told to give their best effort were used for comparison in this study. After being assigned to a group, participants completed a demographics form and were administered the NAART-35, Digit Span, Verbal Fluency, TMT, CVLT-II, TOMM, and Dot Counting Test. Participants were subsequently debriefed with the conclusion of the study occurring approximately one hour after the start. This study looked to determine if normally functioning individuals could be coached to suppress their cognitive functioning without triggering PVTs.

Individuals in Group 3 were consented into the study and answered a series of trivia questions. This study sought to determine if moral elevation can increase a participant's equitable resource distribution. After completing this exercise, the participants filled out a series of questionnaires: Reading the Mind in the Eyes, Moral Foundations Questionnaire, demographics form, 10-Item Five Factor Inventory, Short Dark Triad, Morality Scale of Gender-Roles Attitudes, Beliefs, and Principles Questionnaires, and The Depression, Anxiety, and Stress Scales. Participants were then debriefed with the study taking approximately one hour to complete.

Finally, members of Group 4 were consented into the study and participated in a variety of activities. They filled out a demographics form, completed the SHIPLEY-1, and participated in The Trust Game. They also were administered the Raven's Progressive Matrices Shortened Form which is the portion important to the current study. The purpose of the study was to determine if the pronouns an individual uses impacts how trustworthy they are perceived to be by others. The study took approximately one hour to complete, and participants were debriefed upon exit.

Materials

Televideoconferencing. For the videoconferencing condition, the HIPAA compliant streaming service Zoom was used. Zoom employs a multi-layer security system with Advanced Encryption Standard using 256-bit keys to protect videoconferencing (Zoom Video Communications, Inc., 2019). It provided a high-quality audio and high-definition video through its streaming service. Prior to participating in the study, participants signed up through SONA where an advertisement indicated they must have access to a computer with a video camera and microphone capabilities. Individuals then completed a demographic screener. Participants remained in their own homes, which differed from most studies looking at teleadministration; this set-up provided context for the feasibility of administering teleneuropsychological assessments while individuals remain in their homes and when distractions cannot be entirely controlled for. Participants were seated and viewed the examiner on their monitor; the examiner viewed participants from a similar point of view. The camera in the examiner’s room was adjusted so that the participant was able to view the examiner. Similarly, the participant’s camera was adjusted so the examiner could observe the participant and their visible environment. Assessments that required visual stimuli were provided through the screen-share function of Zoom so that all materials were presented equally to all participants.

Measures

Cognitive Instruments.

North American Adult Reading Test. The North American Adult Reading Test (NAART; Blair & Spreen, 1989) is an assessment that estimates verbal intellectual ability. Individuals are presented a page with 61 irregularly pronounced words printed on it and are instructed to read the page aloud to the best of their ability. Examiners tally the number of corrected read words which results in a raw score that is translated into a standard score. Administration is completed in under 15 minutes and can be markedly less time depending on the response rate of the participant. A full-scale IQ estimate is determined upon the completion of the assessment.

The NAART has an internal consistency of α = .935 and correlations with the Wechsler Adult Intelligence Scale – Revised were good (WAIS-R; VIQ = .83, FSIQ = .75; Blair and Spreen, 1989). Similar results were reproduced by Uttl (2002) as Cronbach's α ranged from .92 - .94 and interrater reliability was .93. When compared to the WAIS-R Vocabulary subtest in Uttl's subsequent study, the validity coefficient was .75 (2002). Test-retest reliability was reported to be .98 over a ten-day period (Crawford, Stewart, Besson, et al., 1989) and one-year retest of .89 (Deary, Whalley, & Crawford., 2004). The NAART was shown to significantly correlate with Rey Verbal Learning Test (.17 - .22; Knight, McMahon, Green, & Skeaff, 2010) and the Paired Associate Test (.23 - .26; Uttl, Graf, & Richter, 2002). This assessment has also been shown to predict performance on Verbal Fluency (r = .41 - .67; Crawford, Moore, & Cameron, 1992; Ross et al., 2003) and Raven's Matrices (r = .59 - .64; Freeman & Godfrey, 2000). At least one study used the NAART while comparing videoconferencing administration and face-to-face; no significant differences were found (Kirkwood, Peck, & Bennie, 2000).

Oral Trails A & B. The Oral Trail Making Test was developed by Ricker and Axelrod (1994) after there were concerns related to physical or visual limitations contributing to performance on the original Trail Making Test (TMT; Reitan, 1958). In this task, individuals are required to count from 1 to 25 as quickly as they can and the examiner times the participant, this constitutes as Oral Trails A. On the Oral Trails B trial, individuals are instructed to switch between numbers and letters in sequential order until directed to stop by the examiner (Ricker & Axelrod, 1994). Oral Trails A & B mirror the practice of the original TMT where participants are asked to draw a line between numbers or letters in the specified order as if to connect the dots (Reitan, 1958). If a participant makes an error, the examiner redirects the participant to their last correct response; timing is continued through corrections (Mrazik, Millis, & Drane, 2010). Time of completion is recorded and used to produce a z-score when compared to norms. It should be noted that while oral and written TMT evaluate the same cognitive paradigm, they are two separate tests with different normed data.

When compared to the written TMT, Oral Trails A correlated at $r = .68$ and Oral Trails B correlated with $r = .72$ (Ricker & Axelrod, 1994). It should be noted that subsequent studies produced weak to mild correlations between the written and oral versions of Trails A ($r = .10$ - .43; Axelrod & Lamberty, 2006; Kowalczyk et al., 2001; Mrazik et al., 2010; Oliveira-Souza et al., 2000). It is speculated that the task of counting to the number of 25 is so simple and engrained that it is not strongly correlated to its written counterpart (Strauss, Sherman, & Spreen, 2006; Kaemmerer & Riodan, 2016). However, results comparing both versions of Trails B indicated larger correlations ($r = .59$ - .62; Mrazik et al., 2010; Oliveira-Souza et al., 2000). Ruchinskas (2003) found that individuals with lower education and lower Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) scores were associated with poorer performance on Oral Trails B. Kowalczyk et al. (2001) showed Oral Trail Making Test moderately correlated with other measures of cognitive flexibility (Digit Symbol, r = -.294 to -.511 and COWAT, r = -.256 to -.434). Orals Trail Making Test also produced strong correlations with the Clock Drawing Test (r = -.55 to -.60) and Digit Span Backwards (-.19 to -.23) (Bastug et al., 2013).

A study completed by Wadsworth et al., (2016) showed participants performed at a comparable rate when administered Oral Trail Making Test though a videoconferencing format as compared to a face-to-face format (ICC = .79 - .83, $p < .001$). It should be noted that there was a significant difference, however, on the average time it took participants to complete Oral Trails A by an average of two seconds as compared to face-to-face testing ($p < .001$); there was no notable difference in Oral Trails B ($p = .726$) (Wadsworth et al., 2016).

California Verbal Learning Test – 2nd Edition. The California Verbal Learning Test – Second Edition (CVLT-II) uses a multiple-trail learning paradigm to assess verbal learning and memory (Delis, Kramer, Kaplan, & Ober, 2000). Participants are provided a list of sixteen words (List A) that belong to four semantic categories; this list is verbally presented for five trials. The participant is asked to repeat any words that they remember, and the examiner records their answers in order. The examiner then reads an interference list consisting of a different sixteen words before asking the participant to recall List A after a short delay. The administration continues with a short-delay cued-recall, a long-delay free- and cued-recall, and recognition questions (Strauss et al., 2006). The CVLT-II takes approximately 20 minutes to administer, 50 total minutes are needed though to account for the delays between trials. Scoring provides T-scores and z-scores when comparing performance to norms.

During test construction, internal consistency was measured through various methods to account for limitations in learning and recall capacity; results ultimately were found to be in the high range (r = .79 - .96; Delis et al., 2000; Kreutzer, DeLuca, & Caplan, 2011). Reliability of scores showed high coefficients on primary measures (r = .80 - .94; Delis et al., 2000; Woods et al., 2006). Test-retest correlation coefficients ranged from .57 - .69 (Alioto et al., 2017). CVLT-II is shown to correlate with Word Memory Test (WMT; Green, 2003) with reported r = .38 - .81 (Armistead-Jehle, Green, Gervais, & Hungerford, 2015; Davis & Wall, 2014) and is positively correlated with the Verbal Paired Associates (VPA; r = .51 - .54; Miller et al., 2012).

Verbal learning and memory are often assessed in neuropsychological assessments. CVLT-II, amongst other measures of verbal learning (Hopkins Verbal Learning Test and Rey Auditory Verbal Learning Test), have been used in teleneuropsychological studies with minimal to no differences in administration format (Cullum et al., 2014; Wadsworth et al., 2018). In a remote assessment of individuals with multiple sclerosis, there was no difference in performance for individuals that tested using CVLT-II through teleservices versus those that attended face-to-face testing ($p > .50$; Barcellos et al., 2017).

Verbal Fluency. Verbal fluency tasks assess an individual's ability to spontaneously produce words with designated parameters (Cohen & Stanczak, 2000; Strauss et al., 2006; Thurstone, 1938). Numerous variations of this task have been established, but the task is often separated into phonemic/letter (Benton, 1968) and semantic/category fluency tests (Newcombe, 1969). Participants are given one minute to produce as many unique words that meet the criteria established by the instructions with the total number of independent words tallied (Shao, Janse, Visser, & Meyer, 2014). Fluency tasks are used to assess executive functioning (Miyake & Friedman, 2012; Takacs et al., 2013), fluid intelligence (Roca et al., 2012), and verbal ability

(Bragrad et al., 2012; Sauzeon et al., 2011). For this study, we used the Controlled Oral Word Association (COWA; Benton & Hamsher, 1976). The comparison group for this study only assessed F phonemic fluency. Total administration for phonemic and semantic fluency is approximately five minutes.

Phonemic Fluency. Phonemic tasks require participants to produce as many words that begin with a particular letter within a set period of time, often one minute. Letters have varied in different studies, but the most commonly used letters are F, A, and S (Strauss et al., 2006). Participants in this study were asked to provide as many words as possible for three letters individually (F, A. and S) in one minute (Barry, Bates, & Labouvie, 2008). Internal reliability (r = .83) and test-retest reliability (r = .74) was found to be high (Tombaugh, Kozak, & Rees, 1999). Ross (2003) found that interrater reliability was high (r = .99). Reading levels (r = .47; Johnson-Selfridge et al., 1998) and WAIS-R FSIQ (r = .44 - .87; Henry & Crawford, 2004; Lacy et al., 1996) are moderately correlated with phonemic fluency. Additionally, Whiteside et al., (2015) found FAS Fluency significantly correlated with WAIS-III Vocabulary subtest (r = .42) and Boston Naming Test (r = .39 - .50; Henry et al., 2004). Phonemic clusters correlated negatively with Stroop Interference score (r = -.25; Ross, et al., 2007).

Letter fluency has been a staple in teleneuropsychological assessments and has been used in over seven studies (Marra, Hamlet, & UF Neuropsychology, 2020a). In comparing participant performance between face-to-face administration and videoconferencing, intraclass correlates ranged from .83 - .93 (Cullum, 2006; Cullum 2014; Wadsworth, 2016) with no significant differences between modalities (p = .188 - .092; Vestal, 2006; Wadsworth 2016; Wadsworth, 2018)

Semantic Fluency. Individuals are asked to articulate as many different words within a singular category during a semantic fluency task. The present study opted for the most commonly used category, animals, and instructed participants to say as many unique animals as they could within one minute (Strauss et al., 2006). At one-week retest, reliability was ranged from .56 - .70 in numerous studies (Bird et al., 2004; Levine et al., 2004; Ross, 2003). Interrater reliability was acceptable (r = .70 - .90; Ross et al., 2007). Semantic fluency predicted semantic clustering on CVLT (r = .21; Abwender et al., 2001). Animal fluency also correlated significantly with Boston Naming Test (r = .37 - .68) and WAIS-III Vocabulary subtest (r = .45) (Henry et al., 2004; Whiteside et al., 2015).

As compared to letter fluency, category fluency had a greater discrepancy in performance across administration formats (Marra et al., 2020a). Intraclass correlates ranged from fair to good (ICC = .58 - .74; Cullum, 2006; Cullum, 2014; Wadsworth, 2016). Wadsworth (2018) did find that there was a small but significant effect regarding testing modality, in favor of in-person testing (p < .001).

Raven's Progressive Matrices. Raven's Progressive Matrices (RPM) utilizes visual assessment to determine an individual's inductive reasoning (McCallum, Bracken, & Wasserman, 2000; Raven, 1938). The test consists of sixty questions broken into five sets, in which participants are asked to choose which item in a multiple-choice selection best completes the pattern displayed (Strauss et al., 2006; Raven et al., 2000). An individual's total correct answers are tallied and the score it then converted to a percentile (Raven et al., 1998). Raven's Progressive Matrices take approximately 40 minutes to administer as there is no specified time limit (Strauss et al., 2006).

For this study, an abbreviated nine-item form will be used (Bilker et al., 2012). Two forms were established with the respective correlations to the original RPM (r = .978-.984; Bilker et al., 2012). As noted by the study, Cronbach's alpha was r = .80-.83. By utilizing this shortened form, the administration time is decreased by 76-82 percent (Bilker et al., 2012). There is limited research utilizing version of RPM, but this study allows for another analysis of this shortened form.

Test-retest reliability and split-half reliability were found to be high (>.80; Burke, 1985; Matthews, 1988; Raven et al., 2000). RPM has displayed moderately strong correlations with other measures of intelligence, such as the Test of Nonverbal Intelligence (TONI), WAIS-R FSIQ, and Shipley Institute of Living Scale (r = .50 - .86; Bostantiopoulou et al., 2001; Burke, 1985; Deary et al., 2004). Additionally, Raven's Progressive Matrices correlates well with Block Design (r = .74, Matthews, 1998; Mills, Ablard, & Brody, 1993).

Digit Span. The Digit Span subtest from the Wechsler Adult Intelligence Test – Third Edition is a measure of attention and working memory (The Psychological Corporation, 1997). For Digit Span Forward, participants are instructed to listen carefully to the examiner as they say numerals; the individual is then asked to repeat the numbers in the same order the examiner said them (e.g., "1-2-4", "1-2-4"). This is continued with every two trials adding an additional numeral to extend the length of the series. The task is completed when the participant misses both trials of the same serial length. Digit Span Backwards follows a similar methodology, although the participant is instructed to say the numeral series backwards (e.g., "2-5-3", "3-5-2"). The total correctly repeated stimuli for both trials individually and when totaled together are translated into scaled scores. The administration takes approximately five minutes to complete.

Stability of test-retest of total Digit Span is considered very good as it is >.80 and internal reliability was also high .80 - .89 (The Psychological Corporation, 1997). If Digit Span Forward and Backward are separated, test-retest falls to the adequate range (.70-.79; Strauss et al., 2006). The Digit Span subtest correlates with Kaplan Baycrest Neurocognitive Assessment (KBNA) Attention/Concentration Index at r = .76 (Leach et al., 2000), Working Memory Index at r = .86 (The Psychological Corporation, 1997), and the Wechsler Memory Scale – III (WMS-III) Working Memory Index at r = .48 - .85 (The Psychological Corporation, 2002). The Brief Test of Attention correlates with Digits Forward (r = .43) and Digits Backward (r = .53) (Schretlen, 1997). Spatial Span from WMS-III correlated moderately with Digit Span (r = .34) (Wilde & Strauss, 2002).

Digit Span Forward, Backward, and Total have frequently been used in teleneuropsychological studies given only auditory stimuli is needed (Marra et al., 2020a). Digit Span Forward has produced intraclass correlations ranging from .590 - .750 (Cullum, 2014; Wadsworth, 2016). There were mixed results in performance regarding administration method for Digit Span Forward; Wadsworth (2018) found no effect in administration modality (p = .276), but in a previous study conducted by them in 2016, there was a small but significant difference favoring face-to-face testing (p = .004). Digit Span Backward produced adequate intraclass correlations (ICC = .545 - .690) and there were no differences in performance across test administration modality (p = .635 - .760) (Cullum, 2014; Wadsworth, 2016; Wadsworth, 2018). When combining the two trials, Digit Span Total has produced good to excellent intraclass correlations (ICC = .720 - .780; Cullum et al., 2006; Grosch, 2015).

Performance Validity Measures.

Test of Memory Malingering. The Test of Memory Malingering (TOMM) is a performance validity measure suited to detect patients feigning memory impairment (Tombaugh, 1996). Participants are shown 50 drawings of common objects for three seconds apiece on each of the two learning trials. They are then provided recognition tasks where a picture they had previously been shown is paired with a novel stimulus, participants are asked to inform the examiner of which picture they had seen during the learning trials (Strauss et al., 2006). If a participant selects a novel picture, the examiner immediately provides corrective feedback. Fifteen minutes after the second learning trial is completed, individuals are presented with a retention recognition booklet where they will again have to select the previously seen stimuli when compared to a novel picture (Tombaugh, 1996). Total correctly identified pictures are tallied out of fifty possible points per trial. Administration takes approximately 15 minutes, not accounting for the 15-minute delay between trial two and retention.

Participants from the community assisted in establishing what was considered a normal performance. On Trial 2, more than 95% of individuals living in the community obtained a 49 or 50 on the second trial. Those that produced suspect effort often scored lower than 45 on Trial 2 or Retention (Tombaugh, 1996; Powell et al., 2004; Teichner & Wagner, 2004). Even when individuals are cognitive impaired, specificity is at 90% with the cutoff established as 45 on Trial 2 and Retention (Rees et al., 1998). Strauss et al., (2006) reported coefficient alphas ranged from .94 - .95 amongst all three trials. When compared to other performance validity measures, TOMM and the Rey 15-Item test were moderately correlated (r = .69-.78; McCaffrey et al., 2003).

A variety of other scales can be derived from the TOMM including the Albany Consistency Index (ACI), the Invalid Forgetting Frequency Index (IFFI), and the TOMMe10. The ACI calculates any inconsistencies across performance of the fifty items over three trials. So should a participant miss any item on the three trials, it is considered an inconsistency (Gunner et al., 2012). However, as this scale does not account for actual learning, the IFFI was established to avoid penalizing the examinee for learning across trials (Buddin Jr., et al., 2014).

Dot Counting Test. The Dot Count Test (DCT; Boone et al., 2002) is a performance validity task that requires an individual to tally the number of dots that are on a notecard sized paper. Participants are instructed to count the dots as quickly as they can without making mistakes. There are twelve stimuli, six are grouped to easily be counted with the remaining six ungrouped and spread around the page. Individuals are assessed for accuracy of their final count per page and speed in which they provide an answer; the combination of these scores produces an Effort Index score (E-Score; Strauss et al., 2006). The administration should take no longer than five minutes to complete.

Specificity was determined to be ≥ 90% when the cut off for the E-score was ≥ 17. Sensitivity was held at 100% for a forensic suspect group and 75% for a civil litigation/disability suspect group (Boone et al., 2002). McCaul et al., (2018) found that if the cut-off was lowered to ≥13.80, sensitivity would increase to 70% while maintaining specificity at ≥ 90%. While assessing individuals with no motive and with seemingly normal effect, over 50% of patients produce an E-score of ≤ 9 (Boone et al., 2002). The DCT is shown to correlate with the b Test (r = .60), Rey 15-Item (r = -.56), Complex Figure (r = .69), and Warrington Recognition Memory (r = .41) (Nelson et al., 2003).

Social Cognition Measures.

Reading the Mind in the Eyes. The Reading the Mind in the Eyes Test (Baron-Cohen et al., 1997) is an assessment that looks to determine a person's ability to theorize what another may be thinking or feeling, and is considered a measure of cognitive empathy. Participants are presented with 36 photographs of the eyes of other people and are asked to select one target word amongst a group of three additional foils that best describe how the person in the picture is feeling (Baron-Cohen et al., 2001). The number of correct interpretations is tallied and this raw total is compared to norms.

Test-retest was found to be good (.63-.83; Fernandez-Abascal et al., 2013; Vellante et al., 2013) and internal reliability has also shown to be adequate (α = .605; Baron-Cohen et al., 2015; Vellante et al., 2013). As expected, Reading the Mind in the Eyes negatively correlated with the Autism Quotient (r = .05) suggesting that those who had a greater Theory of Mind had less qualities associated with autism spectrum disorder (Baron-Cohen et al., 2001). Reading the Mind in the Eyes also correlated with the Vocabulary subtest (r = .49; Peterson & Miller, 2012).

Data Analyses

Missing data.

Prior to addressing the planned hypotheses of the current study, data cleaning procedures were completed and descriptive statistics were assessed in an effort to explore for outliers (i.e., anything beyond three standard deviations from the mean) and to assess the degree of missing data. As all data collected required interactive engagement, with the exception of technological issues, there were minimally missing data from the teleadministration group. Each measure in all groups was examined and it was determined no individual variable had greater than five percent of the data missing. To ensure there were no problems with missing data, Little's

MCAR was conducted and it was determined that the data were missing completely at random (χ^2 1028.47, DF = 1077, *p* = .853). Additional analyses were conducted for each group and each individual measure with no significant findings across all analyses.

Reading the Mind in the Eyes was missing the largest number of items; there were 28 missing variables spanning six participants. A Little's MCAR still determined the data to be missing at random (χ^2 = 214.75, DF = 194, *p* = .146). As Reading the Mind in the Eyes is calculated by summing the total number of items correct, it was determined that if a participant answered at least 60 percent of the items, a sum mean could be entered in place of the missing data. All six individuals that missed items responded to at least 75 percent of questionnaire, thus a mean was used to replace the missing data.

All other assessments were evaluated and there were singular individuals that did not complete F-Fluency, Digit Span, Dot Counting Test, Raven's Progressive Matrices, and Oral Trails B; these individuals were subsequently excluded from analyses utilizing these measures.

Preliminary analyses.

In preparation for meeting the requirements of the proposed analyses, and due to the small number of minority participants, those who identified as Black/African-American (*n* = 50), Hispanic/Latino (*n* = 65), Asian/Pacific Islander (*n* = 67), American Indian, Native Alaskan, Aleutian, or Eskimo (*n* = 5), Bi-racial (*n* = 30), or Other (*n* = 2), were collapsed into a "Non-White" group for analyses. Additionally, as some individuals (*n* = 8) self-identified as genders outside of the binary, these were coded both as all female given the breakdown of the sample's demographics and equally split between males and females for purposes of comparisons amongst groups (Cameron & Stinson, 2019). All assessments, with the exception of the CVLT-II, did not use normed results based on gender (American Psychological Association, 2015; Trittschuh et al., 2018). In following recommendations by the literature, these individuals were scored using both genders and there were no significant differences in their performances thus, individuals were coded as female for the purpose of analyses based on current demographic proportions of collected and comparison groups (Keo-Meier & Fitzgerald, 2017).

Correlational analyses were run to assess for concerns of multicollinearity; no variables were highly correlated with one another unless they were from the same measure. For example, Trial I on the TOMM correlated with the Albany Consistency Score (r = .996), this was taken into account when evaluating the analyses. Similarly, there was a range of correlations across the CVLT-II measure (r = .416 - .916); discretion was used when conducting analyses using these variables so as to not cause multicollinearity problems.

Additionally, tests were conducted to ensure the data met the assumptions of the hypothesis testing analyses. Of the outliers found, we believe the outliers to be legitimate extreme examples of performance and data were not modified. Our analyses were sufficiently powered so as not to be swayed by atypically high or low responses. Given we believed our outliers to be realistic, we did not employ Windsorizing. No modifications to outliers were made.

Analytic Strategy.

Given this study was seeking to affirm the null hypothesis, the literature was consulted to determine the most frequently utilized statistical measures to assess for non-significant difference. Prior studies had the opportunity to conduct limits of agreement analyses because they used the same group re-testing with different test administration modalities (Bland & Altman, 1986; Hildebrand et al., 2004). By using differing groups due to constraints posed by IRB limitations to in-person data collection, this study was unable to conduct these analyses more commonly used including Bland-Altman plots or intraclass correlation coefficients. Thus, analyses were run according to practices by studies that compared at least two separate groups.

When comparing test administration modality, between groups Analysis of Variance (ANOVA) were conducted to determine if administration method effected performance on a variety of neuropsychological measures. Concurrently, a Welch's *F*-test was run as it outperforms a standard ANOVA when the assumption of equal variances is violated (as indicated by a significant Levene's Test for Equality of Variances) and when sample sizes are unequal (Delacre et al., 2019). When the ANOVA was significant, a Welch's modified t-test was run as this does not assume equal variances amongst groups (Welch, 1947). This approximation does not necessitate normal distributions or equal variances, while providing increased control for type one errors (Delacre et al., 2017). Studies with similar designs have used this method for analysis when comparing two separate groups and attempting to determine if test administration modality impacted participant performance (McArdle, 2021; Parks et al., 2021). An effect size was also calculated using eta squared for most variables unless variance was determined to heterogeneous in which an omega squared was reported as it is a more conservative measure of effect (Okada, 2013; Olejnik & Algina, 2003; Yiğit & Mendes, 2018). As there were four groups used to run these analyses, Chi-square and one-way analysis of variances were used to assess group differences. Bonferroni correction was not utilized as the primary goal of this study was to affirm the null hypothesis. When attempting to find non-significance, running this can contribute to the increased likelihood of Type II errors (Perneger, 1998).

The first three hypotheses proposed that there would be no significant differences in performance on neuropsychological testing regardless of test administration modality.

Hypothesis one focused on cognitive measures and multiple one-way Analysis of Variances (ANOVA) were completed to assess if face-to-face or videoconferencing test administration effected performance on cognitive measures. The predictor variable was the test administration modality, and the criterion variables were the individual's performance on North American Adult Reading Test, Oral Trails B (both raw and estimated written times), California Verbal Learning Test – II, Verbal Fluency, Digit Span, and Raven's Progressive Matrices. It should be noted that for analyses including Oral Trails B, Verbal Fluency, Digit Span, and Raven's Progressive Matrices, only 128 participants were included from the teleadministration group due to individuals not completing the testing. As some of the ANOVAs were significant, a Welch's t-test was conducted to account for variance between groups.

Hypothesis two noted there would be no significant differences in performance on performance validity measures between in-person and teleadministration. Multiple ANOVAs were used to determine if test administration modality influenced performance on PVTs. The independent variable was the test administration modality. The dependent variables were the participant's performance on Test of Memory Malingering, Dot Counting Test, Forced Choice on CVLT-II, and Reliable Digit Span. Similarly to the analysis in hypothesis one, only 128 individuals were included in the analyses of Dot Counting Test and Reliable Digit Span due to individuals not completing the assessment in its entirety. In accordance with similar studies and as there were significant differences observed in the ANOVAs, Welch's t-tests were run to determine if test administration modality is significantly different.

The third hypothesis proposed that there would be no observed nor significant difference in performance on social cognition measures when comparing administration modality. This was addressed by a mixed model ANOVA to determine if test administration effected performance on social cognition measures. The predictor variable was modality of test administration, and the criterion variable was the individual's performance on Reading the Mind in the Eyes.

Hypothesis four dictated that NAART-35 would positively correlate with RPM. To address this hypothesis, a linear regression was run to determine if RPM shortened form was an adequate assessment for overall intellectual functioning, regardless of administration modality. The NAART-35 was the criterion variable. The variables were assessed to determine how congruent the measures are in predicting overall intellectual functioning when controlled for modality of administration.

The fifth hypothesis indicated there would be a significant difference in performance between Oral Trails A and Trail Making Test A. To determine this, a between groups ANOVA was used to evaluate any significant differences between testing performances. This expected finding was congruent with the literature on Oral Trails A as it is more of a rote task when asked to verbally count to 25 that when one is asked to visually connect dots with numbers (Strauss, Sherman, & Spreen, 2006; Kaemmerer & Riodan, 2016).

The final and sixth hypothesis suggested RPM and vocabulary would be significant predictors for cognitive empathy. A regression model was used to assess this hypothesis. The independent variables were RPM and vocabulary subtests with the criterion variable as the measured level of cognitive empathy as determined by the Reading the Mind in the Eyes task. Relevant covariates identified in the preliminary analyses were controlled for.

In addition to the prescribed hypotheses, frequencies were run to determine the specificity and sensitivity of the performance validity measures to see if there were any significant differences in performance across administration modality. Similarly, the frequency of individuals scoring at low average or below on cognitive measures was assessed and compared across modalities. As the populations pulled from a college-educated sample, it can be assumed that their cognitive functioning should primarily be within the average range or above (Ritchie & Tucker-Drob, 2018; Strenze, 2007).

CHAPTER FOUR

RESULTS

Baseline Differences in Sociodemographic Information

Across all four groups, 532 individuals enrolled in their respective studies; of those, two individuals were excluded from analyses as they disconnected from their Zoom appointment shortly after consenting to participate in the teleadministration study. After excluding their data, Chi-square and ANOVA assessments were conducted to assess group differences across the sociodemographic characteristics of 530 participants to determine if they could influence results. Refer to Tables 2, 3, and 4 to see results of descriptive statistics and comparisons amongst all four groups.

When Group 1 (teleadministration) was compared to Group 2 (most cognitive/PVT measures) there were no significant differences for the variables gender (p = .492 - .769), race (p = .998), education (p = .382), marital status (p = .066), or number of children (p = .365), but there was a significant difference in age (p = .018). Group 2 restricted the age of their participants to between the ages of 18 and 24, whereas Group 1 did not have that stipulation. Amongst Group 1 participants 16.3 percent of individuals were 25 or older.

Group 1 was compared to Group 3 (Reading the Mind in the Eyes). This analysis showed there were no significant differences for the variables gender (p = .169 - .422), race (p = .398), age (p = .992), education (p = .540), marital status (p = .706), or children (p = .147). It should be noted race was previously significant (p = .002) prior to recoding the two groups, White and Nonwhite. This difference was likely observed because Group 1 had a smaller proportion of non-White individuals (38.0 percent) as compared to Group 3 (42.8 percent). Of particular note, Group 3 had a large Asian/Pacific Islander population (21.7 percent) as compared to Group 1 (5.4 percent).

Finally, Group 1 was compared to Group 4 (Raven's Progressive Matrices). Chi-square analyses found that there were no significant differences on gender (p = .748 - .887), age (p = .356), race (p = .356), and education (p = .153). Group 4 did not collect data related to the patient's marital status or number of children.

Table 2. Demographic comparison between group 1 and group 2

	Group 1 (*n* = 129)		**Group 2 (*n* = 50)**		**χ^2 or *F* value**	***p* value**
	N (%)	*M* (*SD*)	*N* (%)	*M* (*SD*)		
Sex**					.086	.856
Male	39 (30.2)		14 (28.0)			
Female	82 (63.6)		36 (72.0)			
Transgender	4 (3.1)		--			
Nonbinary	4 (3.1)		--			
Age		21.58 (4.75)		19.92 (1.90)	5.74	.018
Race**					.000	1.000
White/non-Hispanic	80 (62.0)		31 (62.0)			
Black/African-American	17 (13.2)		5 (10.0)			
Hispanic/Latino	16 (12.4)		7 (14.0)			
Asian/Pacific Islander	7 (5.4)		4 (8.0)			
Indigenous	1 (0.8)		--			
Bi-racial	6 (4.7)		3 (6.0)			
Other	2 (1.6)		--			
Years of Education		14.36 (1.80)		14.10 (1.65)	.768	.382
Marital Status					5.43	.066
Never married	116 (89.9)		50 (100.0)			
Married	10 (7.8)		--			
Divorced/Separated	3 (2.3)		--			
Children					5.43	.365
Zero	116 (89.9)		50 (100.0)			
One	3 (2.3)		--			
Two	7 (5.4)		--			
Three	1 (0.8)		--			
Four	1 (0.8)		--			
Five	--		--			
Six	--		--			
Seven	1 (0.8)		--			

** = Recoded to fit the assumption of the analyses

Table 3. Demographic comparison between group 1 and group 3

	Group 1 (n = 129)		Group 3 (n = 180)		χ^2 or F value	p value
	N (%)	M (SD)	N (%)	M (SD)		
Sex**					1.89	.184
Male	39 (30.2)		68 (37.8)			
Female	82 (63.6)		112 (62.2)			
Transgender	4 (3.1)		--			
Nonbinary	4 (3.1)		--			
Age		21.58 (4.75)		21.58 (5.45)	.000	.992
Race**					.715	.414
White/non-Hispanic	80 (62.0)		103 (57.2)			
Black/African-American	17 (13.2)		12 (6.7)			
Hispanic/Latino	16 (12.4)		16 (8.9)			
Asian/Pacific Islander	7 (5.4)		39 (21.7)			
Indigenous	1 (0.8)		3 (1.7)			
Bi-racial	6 (4.7)		7 (3.9)			
Other	2 (1.6)		--			
Years of Education		14.36 (1.80)		13.51 (3.90)	.376	.540
Marital Status					.695	.706
Never married	116 (89.9)		157 (87.2)			
Married	10 (7.8)		19 (10.6)			
Divorced/Separated	3 (2.3)		4 (2.2)			
Children					9.503	.147
Zero	116 (89.9)		162 (90.0)			
One	3 (2.3)		8 (4.4)			
Two	7 (5.4)		2 (1.1)			
Three	1 (0.8)		5 (2.8)			
Four	1 (0.8)		2 (1.1)			
Five	--		--			
Six	--		1 (0.6)			
Seven	1 (0.8)		--			

** = Recoded to fit the assumption of the analyses

Table 4. Demographic comparison between group 1 and group 4

	Group 1 (n = 129)		**Group 4 (n = 171)**		**χ^2 or F value**	**p value**
	N (%)	*M* (*SD*)	*N* (%)	*M* (*SD*)		
Sex**					.020	.900
Male	39 (30.2)		53 (31.0)			
Female	82 (63.6)		117 (68.4)			
Transgender	4 (3.1)		1 (0.6)			
Nonbinary	4 (3.1)					
Age		21.58 (4.75)		21.05 (5.02)	.855	.356
Race**					.851	.407
White/non-Hispanic	80 (62.0)		97 (56.7)			
Black/African-American	17 (13.2)		16 (9.4)			
Hispanic/Latino	16 (12.4)		26 (15.2)			
Asian/Pacific Islander	7 (5.4)		17 (9.9)			
Indigenous	1 (0.8)		1 (0.6)			
Bi-racial	6 (4.7)		14 (8.2)			
Other	2 (1.6)		--			
Years of Education		14.36 (1.80)		14.03 (2.10)	2.052	.153

** = Recoded to fit the assumption of the analyses

Hypothesis One: Cognitive Measures

A set of one-way between-groups analyses of variance were conducted to evaluate the impact of administering neurocognitive measures over videoconferencing and to determine if there were any significant differences in test administration modality when compared to in-person testing. Results of ANOVAs run for each of the variables including the NAART, Oral Trail Making Test B, CVLT-II, Digit Span, F-Fluency, and Raven's Progressive Matrices found some significant findings when assessing test administration modality; additionally, Welch's *t*-test and effect sizes were calculated when significance was found. Please refer to Table 5 for full findings.

For Oral Trail Making Test B, two scores were compared to a written, in-person, Trails B Making Test; the first was raw time and the second was the raw time for oral trails converted to an estimated written time. Oral Trail Making Test B (when compared to Written Trail Making Test B) was significant for raw time, $F(1, 176) = 8.46$, $p = .004$. When the raw oral trials time was converted to an expected written score using the formula suggested by Axelrod & Lamberty (2012), $F(1, 176) = 81.16$, $p < .001$, it was also found to be significant. When comparing the means using an independent sample *t*-test, Oral Trail Making Test B was found to be significantly faster than written Trails B, $t(176) = -2.91$, $p = .004$. The magnitude of the differences in the means was small (eta squared = 0.05). When Oral Trails B raw time was converted to an estimated written score, Levene's Test for Equal Variances was found to be significant, thus equal variances could not be assumed. A significant difference was still found when tests were conducted that accounted for unequal variances, $t(172.14) = 9.01$, $p < .001$. The magnitude of the differences in the means was exceptionally large (omega squared = 0.31).

The scores for the CVLT-II were significantly different between groups across all learning trials, list B learning, all recall trials, and recognition hits (please refer to Table 5 for results of all ANOVAs). Of note, there was a significant difference in raw total of all five learning trials (each trial and combined all five trials were $p < .001$), list B learning ($p = .007$), short-delay free recall ($p < .001$), long-delay free recall ($p < .001$), and recognition hits ($p = .003$). The only nonsignificant score was recognition false-positives ($p = .222$). When comparing means, the teleadministration group consistently learned, recalled, and recognized less verbal information than the face-to-face group. Those in the teleadministration group also endorsed more false positives than those in face-to-face testing. Respectively, significance was found across learning ($p < .001$), recall ($p < .001$), and recognition hits ($p = .003$). No significance was found on recognition false positives, $p = .222$. The magnitude of the differences in the means across combined learning trials was large (eta squared = 0.16), long-delayed free recall was large (eta squared = 0.14), and recognition hits was small (eta squared = 0.05). Refer to Table 5 for all reported findings on the CVLT-II.

When evaluating F-fluency, there was a significant difference found across verbal production, $F(1, 176) = 32.13$, $p < .001$. Comparing means, the group that was administered F-fluency through videoconferencing spontaneously produced less words than the group that was administered testing face-to-face. This was found to be significant, $t(176) = -5.67$, $p < .001$. The magnitude of the differences in the means was found to be large (eta squared = 0.15).

The last significant finding on the cognitive measures occurred when comparing Raven's Progressive Matrices, $F(1, 297) = 103.35$, $p < .001$. Further comparison of the means suggested that individuals who were administered this matrices task through teleadministration ($M = 4.55$, $SD = 1.78$) responded with fewer correct answers than the in-person administration ($M = 6.61$, *SD* = 1.70). This was found to be significant, $t(297) = -10.17$, $p < .001$. The magnitude of this difference in means (mean difference = -2.06, 95% CI: -246 to -.166) was exceptionally large (eta squared = .26).

ANOVAs evaluating the impact of test administration modality on the NAART Predicted IQ, $F(1, 177) = .015$, $p = .901$, and Digit Span raw total, $F(1, 176) = 1.39$, $p = .241$, did not have significant findings, suggesting the administration of both of these measures were not impacted by administration modality.

All of the above analyses were re-run removing all participants that had reported technological problems or observed distractions ($N = 18$), and the analyses found the same variables differed significantly across testing modalities. Thus, removing these participants did not impact any of the results.

Table 5. Hypothesis one results.

		N	Mean	*SD*	*F*	*F* (*p*)	*t*	*t* (*p*)	Mean Difference	95% CI lower bound	95% CI upper bound	Eta^2
NAART35 Predicted IQ Score	Tele	129	105.37	6.27	.015	.901						
	F2F	50	105.50	6.58								
Trail Making Test B - Raw Time	Tele	128	40.10	22.16	8.46	.004	-2.91	.004	-10.39	-17.44	-3.34	0.05
	F2F	50	50.49	19.36								
Trail Making Test B - Expected Written*	Tele	128	104.14	59.83	81.16	<.001	9.01	<.001	53.65	41.90	65.40	0.31
	F2F	50	50.49	19.36								
CVLT-II Trial 1	Tele	129	6.19	2.02	18.35	<.001	-4.28	<.001	-1.51	-2.20	-0.81	0.09
	F2F	50	7.70	2.34								
CVLT-II Trial 2	Tele	129	9.11	2.26	34.70	<.001	-5.89	<.001	-2.27	-3.03	-1.51	0.16
	F2F	50	11.38	2.46								
CVLT-II Trial 3	Tele	129	10.84	2.42	23.97	<.001	-4.90	<.001	-1.96	-2.74	-1.17	0.12
	F2F	50	12.80	2.35								
CVLT-II Trial 4	Tele	129	11.79	2.30	20.51	<.001	-4.53	<.001	-1.73	-2.48	-0.98	0.10
	F2F	50	13.52	2.28								
CVLT-II Trial 5*	Tele	129	12.32	2.33	25.66	<.001	-5.07	<.001	-1.74	-2.42	-1.06	0.12
	F2F	50	14.06	1.95								
CVLT-II Trials 1-5 Free Recall Total	Tele	129	50.26	9.31	34.92	<.001	-5.91	<.001	-9.50	-12.28	-6.13	0.16
	F2F	50	59.46	9.46								
CVLT-II List B	Tele	129	5.91	1.83	7.38	.007	-2.72	.007	-0.87	-1.49	-0.24	0.04
	F2F	50	6.78	2.10								
CVLT-II Short-Delay Free Recall	Tele	129	10.45	2.73	34.49	<.001	-5.87	<.001	-2.63	-3.51	-1.75	0.18
	F2F	50	13.08	2.59								
CVLT-II Short-Delay Cued Recall	Tele	129	11.47	2.72	17.68	<.001	-4.21	<.001	-1.84	-2.70	-0.97	0.09
	F2F	50	13.30	2.35								
CVLT-II Long-Delay Free Recall	Tele	129	11.01	2.73	29.63	<.001	-5.44	<.001	-2.43	-3.31	-1.55	0.14
	F2F	50	13.44	2.56								
CVLT-II Long-Delay Cued Recall	Tele	129	11.74	2.62	26.72	<.001	-5.17	<.001	-2.18	-3.01	-1.35	0.13
	F2F	50	13.92	2.27								
CVLT-II Recognition Hits	Tele	129	14.85	1.31	9.07	.003	-3.01	.003	-0.63	-1.04	-0.22	0.05
	F2F	50	15.48	1.07								
CVLT-II False-Positives	Tele	129	2.07	2.76	1.50	.222						
	F2F	50	1.50	2.88								
F-Fluency Raw Total	Tele	128	10.59	3.71	32.13	<.001	-5.67	<.001	-3.73	-5.03	-2.43	0.15
	F2F	50	14.32	4.51								
Digit Span Raw Total	Tele	128	16.97	3.84	1.39	.241						
	F2F	50	17.74	4.15								
Ravens Total Correct	Tele	128	4.55	1.78	103.35	<.001	-10.17	<.001	-2.06	-2.46	-1.66	0.26
	F2F	171	6.61	1.70								

* = Levene's statistic was significant, thus Welch's scores and Omega Squared are reported

Hypothesis Two: Performance Validity Measures

A set of one-way between-groups analysis of variance was conducted to evaluate whether administering performance validity tests over videoconferencing would impact a person's performance and would there be any significant differences when compared to face-to-face administration. It was thought that no significant differences would occur between teleadministration and in-person administration of performance validity measures. Results of individual ANOVAs that analyzed the TOMM, Dot Counting Test, Forced Choice Percent Accuracy on CVLT-II, and Reliable Digit Span found some significant findings between test administration modality. If Levene's test suggested heterogeneity of variances, a Welch's *F*-test was reported. Additionally, Welch's *t*-test and effect size were calculated when significance was found. Please refer to Table 6 for full findings.

The TOMM was found to be significant on Trial 1 raw score and on the Albany Consistency Index (ACI). Trial 1 was significant for total score (Levene's, $p = .007$). Given that the groups' variances were found to be heterogeneous, the Welch's *F*-test was consulted and Trial 1 was still found to be significant, $F(1, 125.44) = 6.45$, $p = .012$. Similarly, the Albany Consistency Index (ACI) was found to be significant for total score (as was Levene's, $p = .003$), $F(1, 133.26) = 6.90$, $p = .010$. Equal variances could not be assumed when comparing means, teleadministration of TOMM Trial 1 and calculated ACI were found to significantly differ from in-person administration of Trial 1 and determined ACI. Both were found to be significant, suggesting that individuals who were administered the TOMM Trial 1, $t(125.44) = -2.54$, $p = .012$, and their subsequently totaled ACI, $t(133.26) = -2.63$, $p = .010$, over teleadministration performed significantly worse than those who completed the task face-to-face. The magnitude of the differences in the means (equal variances not assumed) was small (omega squared = 0.03) for

Trial 1 and ACI (omega squared = 0.03). Differences between administration modality for Trial 2, Retention, Invalid Forgetting Frequency Index, and another indicator of validity, TOMMe10, which tallies the number of errors within the first ten trials on Trial 1, were all found to be insignificant (p = .054 - .334).

Additionally, the CVLT-II's forced-choice recognition was found to be significant. Levene's test suggested that the homogeneity of variances was violated (p = .001) and the Welch's F-test was also significant, $F(1, 167.40) = 4.13$, $p = .044$. Upon further investigation of the means, those who engaged in testing through videoconferencing (M = 15.87, SD = 0.40) were more likely to endorse a novel word when presented with a word pair including a word they had been told previously than compared to those that completed testing in-person (M = 15.96, SD = 0.20), $t(167.40) = -2.03$, $p = .044$. Furthermore, the magnitude of this difference in means (mean difference = -0.09, 95% CI: -0.18 to -0.00) was small (omega squared = 0.02).

An analysis of the Dot Counting Test's E-score was also conducted and found to be significant, $F(1, 176) = 41.92$, $p < .001$. An analysis of means uncovered that those who participated in the teleadministration (M = 12.43, SD = 3.35) performed worse (slower and/or with more errors) than those who engaged in face-to-face testing (M = 8.96, SD = 2.82). This was again found to be significant, $t(176) = 6.48$, $p < .001$. The magnitude of the differences in means (mean difference = 3.47, 95% CI: 2.41-4.53) was large (eta squared = 0.19).

Reliable Digit Span was found to be non-significant (p = .134) suggesting that there were no meaningful differences in performance across groups, regardless of test administration modality.

Even when removing all participants that had reported technological problems or observed distractions (N = 18), re-running the analyses found the same variables differed significantly across testing modalities, with the exception of the CVLT-II Forced Choice Recognition Accuracy which was no longer significant (p = .125).

Table 6. Hypothesis two results.

		N	Mean	*SD*	*F*	*F* (*p*)	*t*	*t* (*p*)	Mean Difference	95% CI lower bound	95% CI upper bound	Eta^2
TOMM Trial 1 Score*	Tele	129	47.93	2.51	6.45	.012	-2.54	.012	-0.85	-1.51	-0.19	0.03
	F2F	50	48.78	1.78								
TOMM Trial 2 Score**	Tele	129	49.93	0.34	2.16	.143						
	F2F	50	50.00	0.00								
TOMM Retention Trial Score	Tele	129	49.95	0.21	.669	.415						
	F2F	50	49.98	0.14								
Albany Consistency Index*	Tele	129	47.86	2.66	6.90	.010	-2.63	.010	-0.90	-1.58	-0.22	0.03
	F2F	50	48.76	1.77								
Invalid Forgetting Frequency Index*	Tele	129	49.91	0.31	171.91	.054						
	F2F	50	49.98	0.14								
TOMMe10	Tele	129	9.75	0.60	.841	.360						
	F2F	50	9.84	0.51								
DCT E-Score	Tele	128	12.43	3.35	41.92	<.001	6.48	<.001	3.47	2.41	4.53	0.19
	F2F	50	8.96	2.82								
CVLT-II Forced-Choice Recognition Accuracy*	Tele	129	15.87	2.52	4.128	.044	-2.03	.044	-0.09	-0.18	-0.00	0.02
	F2F	50	15.96	1.24								
Reliable Digit Span	Tele	128	9.73	2.19	2.27	.134						
	F2F	50	10.30	2.40								

* = Levene's statistic was significant, thus Welch's scores and Omega Squared are reported

** = there was no variance in one of the groups so a Welch's *f*-test could not be calculated, the scores here should be interpreted with caution as the variances are unequal

Hypothesis Three: Social Cognition Measures

A one-way between-groups analysis of variance was conducted to evaluate if test administration modality would influence a person's performance on a measure of social cognition. It was thought that no significant findings would occur between teleadministration and in-person administration. Results of the ANOVA assessing this with the Reading the Mind in the Eyes test found no significant findings when assessing test administration modality, $F(1, 307) = .023$, $p = .880$. These findings indicate that there were no significant differences in performance on this measure of social cognition, regardless of test administration modality. All participants that had reported technological problems or observed distractions ($N = 18$) were removed and re-running the analysis again found no significant differences.

Hypothesis Four: Positive Correlation Between Measures of Intellectual Abilities

A correlation was run between the NAART and the shortened form of Raven's Progressive Matrices as it was hypothesized that these two measures would be positively related, because it is thought that both measures are indications of overall intellectual abilities. The analyses revealed no association between the NAART and Raven's Progressive Matrices, $r = .131$, $p = .151$. This suggested that continued research on the shortened form of Raven's Progressive Matrices (with nine items) may be useful in determining its utility in its shortened form and whether the shortened form provides an effective estimate of overall intellectual functioning. After removing all participants that reported technological problems or observed distractions ($N = 18$), re-running the analysis produced the same insignificant results.

Hypothesis Five: Trails A Paradigms

A one-way ANOVA was conducted to determine if there was a significant difference in completion time when comparing Oral Trail Making Test A to the written Trail Making Test A.

As supported by literature, it was believed that there would be a significant difference, and this was replicated in the current study. As Levene's test suggested heterogeneity of variances, therefore a Welch's *F*-test was reported. Additionally, Welch's *t*-test and effect size were calculated when significance was found. Please refer to Table 7 for details of the analyses.

Raw completion time of Oral Trail Making Test A (when compared to Written Trail Making Test A) found that homogeneity of variances could not be assumed ($p < .001$) and was determined to be significant as assessed by Welch's *f*-test, $F(1, 57.73) = 134.57$, $p < .001$. It was also found to be significant when raw time was converted to an estimated written score as outlined by Axelrod & Lamberty (2012), $F(1, 177) = 23.33$, $p < .001$. Comparison of means, with equal variances not assumed, for raw time indicated Oral Trail Making Test A ($M = 8.88$, $SD = 3.81$) was found to be significantly faster than written Trails A ($M = 22.64$, $SD = 8.04$), $t(57.73) = -11.60$, $p < .001$. The magnitude of the differences in the means was exceptionally large (omega squared = 0.43). Even when converted to estimated written scores, significance was still found between tests, $t(177) = 4.83$, $p < .001$. After conversion of the teleadministration group ($M = 32.87$, $SD = 14.10$), when compared to in-person testing ($M = 22.64$, $SD = 8.04$), the magnitude of effect size (mean difference = 10.23, 95% CI: 6.05 to 14.41) was moderate (eta squared = 0.12). Like the findings comparing Oral and Written Trails B, there is a significant difference in performance comparing these two tasks. Even when raw times are manipulated using conversion equations, there remains a significant difference across administration modalities.

After removing all participants that had reported technological problems or observed distractions ($N = 18$), re-running the analyses found the Trials A paradigm still differed significantly across testing modalities.

Table 7. Hypothesis five results.

		N	Mean	*SD*	*F*	*F* (*p*)	*t*	*t* (*p*)	Mean Difference	95% CI lower bound	95% CI upper bound	Eta^2
Trail Making Test A – Raw Time*	Tele	129	8.88	3.81	134.57	<.001	-11.60	<.001	-13.75	-16.13	-11.38	0.43
	F2F	50	22.64	8.04								
Trial Making Test A – Estimated Written Score	Tele	129	32.87	14.10	23.34	<.001	4.83	<.001	10.23	6.05	14.41	0.12
	F2F	50	22.64	8.04								

* = Levene’s statistic was significant, thus Welch’s scores and Omega Squared are reported

Hypothesis Six: Predictive Value for Cognitive Empathy

The final and sixth hypothesis suggested RPM and vocabulary as assessed by the NAART would be significant predictors for Reading the Mind in the Eyes. This was investigated using a regression model while controlling for covariates including gender, age, education, and race. A moderate relationship (as suggested by Cohen, 1998) was found suggesting the NAART can reliably predict performance on a measure of cognitive empathy (r = .475, $p < .001$); the Raven's Progressive Matrices was not found to be as significant predictor. These results were reproduced even when all participants that had reported technological problems or observed distractions (N = 18) were removed from the analysis.

Exploratory Analyses

Performance Validity Measures.

After collecting the data and running analyses, it was thought that looking at the frequency of individuals who failed performance validity measures would be clinically useful to evaluate if individuals who were administered performance validity measures over teleneuropsychology produced increased rates of invalid performances. While there were significant differences detailed above, particularly on PVTs, determining if it effected the specificity of the measures would be advantageous to the field. As such, the teleadministration database was recoded to include variables indicating whether an individual passed a PVT or not. Further analyses found that 1.6 percent of individuals in the teleadministration group failed TOMM Trial 1 (scores ≤ 41; Martin et al., 2020), 11.7 percent of individuals failed Dot Counting Test (scores ≥ 17; Schroeder & Martin, 2021), 3.1 percent of individuals failed Reliable Digit Span (scores ≤ 6; Schroeder et al., 2012), and 1.6 percent of individuals failed CVLT-II Forced Choice (scores ≤ 14; Schwartz et al., 2016). No individuals failed Trial 2 or the Retention trial of TOMM (scores ≤ 44 on either trial). Upon determining the validity of performance, 1.6 percent of individuals (*n* = 2) failed two or more PVTs across the four that were administered and were thus considered invalid performances. Given this is such a small number of individuals, assessment of sensitivity was not feasible and was subsequently not run.

Specificity was evaluated in the teleadministration group and it was found that cut-offs were comparable to those established in the research with in-person evaluations. In order to maintain specificity at or above 90 percent within the current teleadministration group, the teleneuropsychology data produced the following cut-offs: a cut-off of ≤ 43 on TOMM Trial 1, a cut-off of ≤ 49 on TOMM Trial 2, a cut-off of ≤ 49 on TOMM Retention, a cut-off of ≥ 18 on DCT E-Score, a cut-off of ≤ 7 on Reliable Digit Span, and a cut-off of ≤ 14 on CVLT-II Forced Choice recognition. When compared to the cut-offs for these stand-alone and embedded PVTs in the literature, most cut-offs produced by the teleneuropsychology group that maintained 90 percent specificity were increased as compared to cut-offs used in clinical practice (Schroeder & Martin, 2021). The exception to this was the DCT E-Score which suggested the cut-off in the community may be too strict and incorrectly label valid performances as invalid when administered over videoconferencing.

General Cognitive Measures.

As there were many significant differences in performance on cognitive measures, it was thought evaluating the frequency of individuals scoring at the low average range (1.0 standard deviations below the normative mean) and exceptionally low range (2.0 standard deviations below the normative mean) could be useful given the sample is a college-educated group that could be expected score in the average range (Ritchie & Tucker-Drob, 2018; Strenze, 2007). Frequency analyses were conducted for Oral Trails, Digit Span, F fluency, Semantic fluency, and

CVLT-II. The NAART, an estimator of premorbid intellect, was not evaluated as no individuals across testing groups scored outside of the average or high average range.

On Oral Trails A, 60.5 percent of individuals scored at or below low average when compared to normative data. Consequently, 51.9 percent of individuals scored within the exceptionally low range, with z-scores ranging from -2.02 to -28.66. Regarding Oral Trails B, 48.1 percent of individuals scored at or below the low average range, with 26.6 percent of individuals falling within the exceptionally low range. Z-scores for those within the exceptionally low range on Oral Trials B ranged from -2.01 to -12.32.

The scaled scores were calculated for Digit Span. 15.6 percent of individuals fell within the low average range with no individuals dropping into the exceptionally low range. These scaled scores ranged from 7 to 5. Z-scores produced for F-Fluency found that 50 percent of individuals within the teleadministration group scored within or below the low average range. Furthermore, 14.1 percent of individuals produced exceptionally low scores with z-scores that ranged from -2.12 to -3.55. On a test of semantic fluency, when compared to a normative sample, only 8.6 percent of individuals fell at or below the low average range, and only 2.3 percent were regarded as exceptionally low with z-scores ranging from -2.02 to -2.39.

On the CVLT-II, a learning t-score, a short-delay free-recall z-score, and a long-delay free-recall z-score were evaluated. Across five learning trials, 27.1 percent of individuals scored low average or below and 6.2 percent of the total sample produced an exceptionally low score. With the short-delay free recall, 38.0 percent of the sample fell within or below the low average range and 10.1 percent producing scores considered exceptionally low. On the long-delay free-recall trial, 45.7 percent of individuals produced at least low average scores with 9.3 percent falling within the exceptionally low range.

A comparison of these frequencies to the data collected in person with a comparable group can be found on Table 8.

Table 8. Exploratory analyses of cognitive performance frequencies

	Group 1 (*n* = 129)	**Group 2 (*n* = 50)**	χ^2
	N (%)	*N* (%)	
Trials A			42.88**
≤ 1.0 SD (Low Average)	78 (60.5)	13 (26.0)	
≤ 2.0 SD (Exceptionally Low)	67 (51.9)	0 (0.0)	
Trials B			17.54**
≤ 1.0 SD (Low Average)	62 (48.1)*	11 (22.0)	
≤ 2.0 SD (Exceptionally Low)	34 (26.6)*	0 (0.0)	
Digit Span			0.07
≤ 1.0 SD (Low Average)	20 (15.6)*	7 (14.0)	
≤ 2.0 SD (Exceptionally Low)	0 (0.0)*	0 (0.0)	
CVLT-II Learning Trials 1-5			8.27**
≤ 1.0 SD (Low Average)	35 (27.1)	4 (8.0)	
≤ 2.0 SD (Exceptionally Low)	8 (6.2)	0 (0.0)	
CVLT-II Short-Delay Free Recall			13.57**
≤ 1.0 SD (Low Average)	49 (38.0)	5 (10.0)	
≤ 2.0 SD (Exceptionally Low)	13 (10.1)	2 (4.0)	
CVLT-II Long-Delay Free Recall			21.48**
≤ 1.0 SD (Low Average)	59 (45.7)	5 (10.0)	
≤ 2.0 SD (Exceptionally Low)	12 (9.3)	3 (6.0)	
F-fluency			
≤ 1.0 SD (Low Average)	64 (50.0)*	12 (24.0)	9.97**
≤ 2.0 SD (Exceptionally Low)	18 (14.1)*	3 (6.0)	
Animals Fluency			--
≤ 1.0 SD (Low Average)	11 (8.6)*	--	
≤ 2.0 SD (Exceptionally Low)	3 (2.3)*	--	

* = Total *n* is equal to 128 for these frequency analyses
** = $p < .05$

CHAPTER FIVE

DISCUSSION

With the recent COVID-19 pandemic highlighting the utility of telehealth services, this study sought to determine the feasibility of administering neuropsychological assessment through videoconferencing. The purpose of this study was to evaluate whether there was a significant difference in performance across groups comparing in-person to tele- administration across cognitive assessments, performance validity testing, and social cognition measures. Furthermore, this study looked to determine if measures with limited validity and reliability data (e.g., Raven's Progressive Matrices, shortened form) were related to other substantiated assessments. Given the results found in the previous chapter, we also performed exploratory analyses of performances across administration modality, particularly in the frequency of lower performances and whether measures maintained a similar specificity.

Cognitive Measures

Hypothesis one proposed that cognitive measures can be administered via teleadministration with no significant differences across administration modalities. The results of our findings did not support this hypothesis, suggesting that there is variability in performance across measures when comparing groups who were administered these measures face-to-face as compared to teleadministration. Some measures, including Oral Trail Making Test B, CVLT-II, F-fluency, and Raven's Progressive Matrices (shortened form) were all found to differ significantly across testing modalities with individuals who received in-person testing performing significantly better; thus, the majority of our findings did not support our primary hypothesis. While the NAART and Digit Span did not show any significant differences, a larger portion of the cognitive measures administered did in fact show significant differences.

Furthermore, the results demonstrated that individuals who were administered the measures through videoconferencing regularly scored lower than a comparable group who received these measures through standard administration. These findings are contrary to a variety of pre-pandemic literature with recent meta-analyses evaluating large quantities of data (Brearly et al., 2017; see Table 1 on page 14 for details regarding this extensive literature) that support videoconferencing as a comparable modality of test administration.

There are various possibilities as to why significance was found across testing administrations. First and foremost, it is important to acknowledge that the data was collected over videoconferencing in the middle of a pandemic, which could account for many of the discrepancies observed. Over the course of the COVID-19 pandemic, individuals experienced large disruptions on a societal level including distanced learning, political unrest, increased unemployment, racial injustices, and forced physical separation from others (United Nations, 2022). While individuals have experienced global impacts, there have also been increased rates of mental illness with estimates suggesting an increase of 25 percent in depression and anxiety worldwide (World Health Organization, 2022). Furthermore, studies have found that psychological distress was higher in women, higher in individuals with more education, and more prominent amongst adults aged 25 to 44 (Patel et al., 2022). Additionally, more individuals are using and/or abusing illicit substances over the course of the pandemic (Farhoudian et al., 2021). Deaths related to substance use have increased by nearly 30 percent since the start of the pandemic, resulting in over 99,000 deaths in the first year of the pandemic (Center for Disease Control and Prevention, 2022). Any or all of these factors could have impacted participants' abilities to engage with testing and subsequently producing lower scores.

In the context of these stressors, this study added the use of videoconferencing which may have contributed itself to the lower performance of this sample.

One argument that could be made is the possibility of "Zoom fatigue" impacting one's ability to attend to information for a period of time. Zoom fatigue is defined as the "somatic and cognitive exhaustion that is caused by the intensive and/or inappropriate use of videoconferencing tools" (Riedl, 2022, p. 157). This study required individuals to attend to their computer screens, participating in tasks that were cognitively demanding for over an hour. It is plausible that the participants could have had decreased engagement given the requirements of attending to a screen for an extended period of time.

Furthermore, another factor that may have contributed to decreased performance in those who received the assessments over videoconferencing is that these individuals remained in their home, an environment that cannot be controlled in the same way as a clinic. Despite the growing literature, very few studies have conducted these evaluations with the participant remaining in their home (Cullum et al., 2019; Cullum et al., 2020; Stolwyk & Hammers, 2020). It is possible that participants who were in their homes were unintentionally preoccupied despite being instructed to place themselves in a quiet location with minimal distractions. Previous research has shown that auditory distractions (such as dogs barking, other residents in the home, a phone ringing, etc.) can impact measures of attention (Lavie, 2010), memory (Wais & Gazzley, 2014), and executive functioning (Beaman et al., 2014). Similarly, visual distractions (such as a television in the background, a phone lighting up, etc.) can also impact measures of attention (Craik et al., 1996) and memory (Fernandes & Moscovitch, 2000; Rae & Perfect, 2014). Despite this evidence, other studies have found these minor distractions described above do not meaningfully impact cognitive performances (Craik, 2014; Rae & Perfect, 2014). Evaluation of how distraction may impact cognitive performances have illustrated that individuals are more easily distracted when completing easier tasks, rather than more engaging or difficult tests (Scheiter et al., 2014), whereas other studies have found the opposite (Britton et al., 1983). Additionally, there were a few studies that had participants report that they felt more engaged with the testing over videoconferencing, as they experienced decreased anxiety (Brearly et al., 2017; Jacobsen et al., 2003; Kirkwood et al., 2000). As the most consistent differences emerged on the CVLT-II, a task that requires the participant to attend to a list of words and be able to recall them later, it is possible that distractions discussed above in addition to other factors that may have contributed to the lowered performances.

As alluded to previously, another possibility that could have contributed to the videoconferencing groups' poorer performance could be related to technical issues such as network speeds, display or camera quality, speaker quality, delays in transmission, and comfortability with videoconferencing. Given that data collection occurred in the height of the COVID-19 pandemic, and it is with a college population who had transitioned to distance learning, it could be assumed that these would have impacted testing minimally. Even when removing all participants that had reported technological problems or observed distractions (N = 18), re-running the analyses found the same variables differed significantly across testing modalities.

Additionally, it is possible that administering neuropsychological measures through videoconferencing impacts the participant's performance and inappropriately dampens their scores. Previous research though (see Table 1) suggests this is not the case and indicates that teleadministration produces comparable results to testing in-person. However, there are a few studies that have found verbal word list and story memory tasks (e.g., HVLT-R and WMS-R

Logical Memory) to have a significant difference between test administration modality (Cullum et al., 2006; Jacobson et al., 2003). Even so, it is important to note that most studies have been able to use the same group and compare their individual performances across administration modality. This study faced constraints implemented by the pandemic and subsequently had to use separate comparison groups which may have also contributed to the significant differences between testing administration method.

As discussed above, despite the literature suggesting comparable performances across most cognitive measures, this current study found many significant differences neurocognitive assessments, thus our first hypothesis was not supported by the results. A variety of factors could have contributed to these starkly different performances across testing modalities including a worldwide pandemic, societal conflicts, zoom fatigue, environmental distractors, technological issues, and using separate groups.

Performance Validity Measures

The second hypothesis stated that no significant differences would be observed on performance validity testing regardless of test administration modality. The results of this study did not support this hypothesis and suggested there was variability in performance across these validity measures when comparing groups who were administered these assessments face-to-face versus over teleadministration. It was determined, while the forced-choice paradigm of CVLT-II and Reliable Digit Span did not differ significantly, there was a significant difference on the TOMM and Dot Counting Test. Although the utility of performance validity measures administered over videoconferencing has not been evaluated explicitly, it could be assumed given the literature on cognitive measures that there would be minimal differences in performance. Innumerous reasons could have contributed to these findings and aided to the significant differences observed.

As discussed before, the impacts of the worldwide pandemic and conflicts in society could have created dampened scores on test. Similarly, Zoom fatigue may have contributed to lower performance. The TOMM takes approximately 200 seconds to administer a singular trial (three seconds for viewing the image, one second for switching the image). This requires the participant to stare attentively at their screen for this allotted time. It is possible that, as data collection occurred in the height of the COVID-19 pandemic, participants were experiencing fatigue, general exhaustion, and difficulties focusing that can accompany extended videoconferencing (Abdelrahman, 2021).

If a participant had increased trouble focusing as a result of the COVID-19 pandemic, then another possibility that could have contributed to variable scores were any present distractions. As mentioned before, this study had participants within their own homes so there was little to no control over distractors in their environment. While it is possible an individual may have been distracted during testing which led to lowered scores, previous studies have shown that distracted individuals perform similarly to control participants on the TOMM and other performance validity measures including the Word Memory Test, Word Choice Test, and the Medical Symptom Validity Test (Barhon et al., 2014; Batt, Shores, & Chekaluk, 2008; Shura et al., 2019). Despite this, these studies were conducted within a controlled environment (e.g., a clinic) and distractions were orchestrated, rather than occurring within the participant's natural environment. These increased familiar distractions and the comfortability of remaining in one's home could have contributed to the significant differences in performance across administration modality.

The significant differences found in the Dot Counting Test also were unanticipated; however, there is good reason to assume that technical difficulties accounted for a portion of this discrepancy. As discussed before, the Dot Counting Test requires individuals to count a series of dots on their screen and then provide the total number aloud. When the screenshare feature is used, it minimizes the faces of the individuals on the videoconference call and shifts these images to the side or top of the screen. While attempts were made to ensure that all dots were visible regardless of device used by participants (prior to data collection), we could not control for the settings on their device, nor did we tell participants that all dots were visible as that was outside of the scope of standardized instructions. This was brought to the lead author's attention when a participant asked a research assistant if the windows that showed the examiner and the participant's faces were covering any dots on the screen. Although it was determined that no dots were covered upon further investigation, this situation could have contributed to other participants having similar concerns or confusion, possibly extending their time/errors in attempts to not miss any dots.

Social Cognition Measures

The third hypothesis indicated that there would be no significant difference in performance on measures of social cognition, regardless of test administration modality. This hypothesis was affirmed during our current study as individuals performed comparably between groups that were administered Reading the Mind in the Eyes. There are no studies known at this time that have evaluated the effectiveness of administering this test through videoconferencing, although the measure has regularly been administered through online means (Baron-Cohen et al., 2015; Olderbak et al., 2015). Currently, the test appears readily available to take online for the layperson in a study that was approved by the Harvard IRB (Lab in the Wild, 2022). As no studies reviewed reported differences in data collection methodology, it is likely that Reading the Mind in the Eyes can be administered across videoconferencing or teleadministration with minimal impact on participant performance.

Insignificant Predictor of Intellectual Functioning

The fourth hypothesis, that Raven's Progressive Matrices (shortened form) would predict performance on the NAART, was not supported by the data. Freeman & Godfrey (2000) found that the NAART would adequately predict the full Raven's Matrices (r = .59 - .64), while additional research suggests Raven's Progressive Matrices is correlated with other measures of intelligence (r = .50 - .86; Bostantiopoulou et al., 2001; Burke, 1985; Deary et al., 2004), these findings were not supported within the current study with the shortened form. There are a variety of factors discussed previously that may have contributed to this incongruency with literature including complications with teleadministration, distractibility, and having participants remain in their homes. In addition to these factors, the shortened form of Raven's Progressive Matrices has had limited research utilizing this form; it could be possible that this form does not adequately represent the full-item test, although Bilker et al., (2012) found strong correlations between the shorten form and the full version (r = .98). The findings of this study indicate that further evaluation of the shortened form may be needed.

Trails A Paradigms

The fifth hypothesis proposed that Oral Trail Making Test A and written Trail Making Test A would be significantly different as it was anticipated that those who engaged in the oral task would perform faster than those on the written task. This hypothesis was supported by the results of the current study. Our results are consistent with the literature as Axelrod & Lamberty (2012) noted there were discrepancies in performance between the two paradigms and

Wadsworth et al., (2016) found similar results when utilizing teleadministration. Given the motor element is removed from the oral task, counting upwards is a rote task that individuals learn, practice, and use from an early age, thus requiring little cognitive effort (Axelrod & Lamberty, 2012). The results of our study supported this finding and our hypothesis was affirmed.

What is increasingly interesting is that when implementing a conversion rate suggested by the creators of the Oral Trail Making Test to have an estimated written time (Axelrod & Lamberty, 2012), the results were still significant. However, the results with the estimated written time suggested an opposite relationship, that those who participated in the teleadministration would have performed slower than compared to individuals who completed face-to-face testing. As such, this discrepancy is important to note because it suggests the conversion formula may need to be adjusted.

Prediction of Cognitive Empathy

The sixth and final hypothesis proposed that Raven's Progressive Matrices and vocabulary as measured by the NAART would adequately predict an individual's score on Reading the Mind in the Eyes, while controlling for covariates. A regression was conducted, and the hypothesis was partially supported by the results of our current study as the NAART was able to predict a measure of cognitive empathy. Raven's Progressive Matrices contributed little to the model, and was not a significant predictor, thus only half of this hypothesis was supported. Peterson & Miller (2012) previously found Reading the Mind in the Eyes to significantly correlate with a vocabulary measure. Additionally, a multiple regression found that RPM was primarily associated with emotion recognition tasks and was partially mediated by a Theory of Mind task (Lee et al., 2016). Again, these findings suggest continued evaluation of the RPM shortened form is needed.

Exploratory Findings

Frequency and Specificity of PVTs.

Exploratory analyses were conducted to assess the frequency of which individuals might produce invalid scores on these measures. In accordance with cut-offs used in the community, the vast majority of individuals produced valid scores. The Dot Counting Test produced the highest number of individuals (11.6 percent) that scored above the recommended cut-off (≥ 17 on E-Score), suggesting they took longer to count the dots and/or they provided an incorrect total number of dots on a trial. Using even more liberal cut-offs (as suggested by An et al., 2016; McCaul et al., 2018; Soble et al., 2018) of ≥14 on E-Score, the members of the teleneuropsychology group had 20.3 percent of individuals fail this performance validity measure.

Similar to the discussion above regarding possible influences that contributed to a higher DCT E-Score, it appears that individuals may have experienced technical difficulties related to the imagery of the test. If they did not, it is possible their concern about not counting all of the dots could have caused them to take longer counting the dots. Clinically, if these performances are interpreted as an indication of invalidity and their increased time was a result of technical issues, this may inappropriately inflate the number of individuals classified as producing an invalid performance. While in the current study, the DCT was the only PVT that suggested higher failure rates, these findings suggest continued need to evaluate validity measures when using videoconferencing as new norms to maintain specificity may be necessary.

Frequency of Lower-than-Expected Cognitive Testing.

Additional frequency analyses were conducted outside of the hypotheses again, to evaluate how many individuals scored below expectations by either one standard deviation or two standard deviations. When compared to the sample collected in person, many more individuals who participated in teleadministration scored within these lower ranges of performance. See Table 8 for frequency comparisons. The performances obtained in the current study do not align with expected results, given that a meta-analysis conducted by Brearly et al., (2017), found that across testing, individuals who were administered assessments through videoconferencing, scored on average 1/10th lower than in-person administration participants which was determined to be insignificant. The current study results suggest that a greater number of individuals scored within the low average to exceptionally low range, which would influence interpretation in clinical settings.

The current demographics of the experimental group suggest that the majority of individuals should be scoring within the average range (Ritchie & Tucker-Drob, 2018; Strenze, 2007). As compared to the in-person testing, Trails A, Trails B, and phonemic fluency all had over 50 percent of individuals scoring lower than expected. CVLT-II approached these numbers with 27.1 to 45.7 percent of individuals performing lower than expected on learning and recall trials. Understanding these potential differences and the impacts of a variety of factors listed above possibly contributing to significant differences is important to document and account for, particularly if evaluations are assessing for neurocognitive disorders and/or are forensic in nature.

Implications

Prior to the COVID-19 pandemic, the field of neuropsychology and the use of pencil and paper assessments had largely been stagnant. While early studies using telephones and videoconferencing were seen in other healthcare settings, few individuals in neuropsychology looked to embrace the growing technology available to clinicians and their patients (Stolwyk &

Hammers, 2020). The recent pandemic placed a spotlight on the limited research that has been conducted in neuropsychology as it relates to providing telehealth services, particularly when administering assessments. For the past three decades, cognitive measures have been increasingly evaluated as to their utility over videoconferencing or telecommunication, but no research to date had appraised if there were differences when administering performance validity measures and social cognition tests across this modality. A significant portion of the research suggests that there is little to no difference across testing administration modalities and that it is feasible to provide cognitive assessment services over teleneuropsychology (Cullum et al., 2014; Hildebrand et al., 2004; Loh et al., 2007; Kirkland et al., 2000; Vestal et al., 2006; Galusha-Glasscock et al., 2015; Wadsworth et al., 2018). Results from these studies are available for review in Table 1. The current study looked to expand the available literature to include measures of social cognition and performance validity tests.

While previously established literature suggests that teleneuropsychological services could extend services to underserved populations, the findings of our current study indicate that additional research and evaluation is needed, particularly if patients are looking to remain in their homes for appointments. Three studies prior to the pandemic had patients participate in evaluation from their home (Abdolahi et al., 2016; Lindauer et al., 2017; Stillerova et al., 2016) and these studies found no significant differences. Others yet opted for a mixed model of the patient completing the intake interview and feedback from home, while visiting the clinic for testing (Stolwyk & Hammers, 2020). Most commonly though, the investigators would have the patient visit onsite in their clinic and provided all services (including assessment) over videoconferencing within the clinic environment (Brearly et al., 2017).

The results of the current study point to the need for additional research with expanded batteries and variable environments that could contribute either positively or negatively to the patient's performance. Although the cognitive measures administered in the current study had been shown to be valid within a telehealth setting, this was not the case across the board in the current investigation. Additionally, there were interesting results in the increased frequency of lower-than-expected performance across cognitive measures and lends to future research questions related to teleneuropsychology and what factors might be contributing to this discrepancy. There are some promising results across the utility of performance validity and social cognition measure through videoconferencing administration.

Should individuals want to implement teleneuropsychological assessment within their practices, the results of this study and the literature suggest a few key points that must be kept in mind when making these determinations. One must consider the capabilities of their patient as technical issues can disrupt data collection. Additionally, access to technology itself may be limited if a patient is from a lower socioeconomic background and should be addressed as reading information on a phone rather than a computer screen may be an opportunity for variable results. While the literature supports numerous cognitive domains that prove to be comparable across testing modalities, most of the research has primarily been conducted in controlled settings rather than in a patient's home. Ultimately, in a meta-analysis (Brearly et al., 2017), most neuropsychological measures were able to be administered over videoconferencing. Moderators in the results included age and internet connection, with individuals older than 75 and that had slower connections resulting in less consistent results. Furthermore, motor testing was too heterogenous for interpretation in the meta-analysis. The literature suggests that the use of neuropsychological testing over videoconferencing is feasible while in a controlled environment (IOPC, 2020), but the results of our study suggest that continued research is needed regarding patients remaining in their own homes.

Ultimately, the utility of teleneuropsychology could impact a large variety of clinicians and patients, by opening up available services to those that are underserved and/or live in a rural area with low numbers of neuropsychologists. The results of the current study indicate that there were significant differences across testing modalities and this would need to be evaluated further. Despite these differences in scores, the vast majority of participants produced scores on performance validity testing that would allow for interpretation and identified specificity was maintained or above 90 percent when using cut-offs for previously established for in-person administration.

Strengths and Limitations

One of the greatest strengths of this study is that at the time of this analysis, there were no studies that had looked at the feasibility of administering performance validity measures and social cognition measures over videoconferencing. Given the variability of results, particularly when compared to the literature, this study provides a counter perspective that suggests additional considerations should be taken when considering performing teleneuropsychological assessment through videoconferencing, particularly when the patients are in their own home. By having participants within their own home for this study, it allowed for an evaluation in an environment that cannot be controlled for, offering different situations than one would experience within a neuropsychology clinic (e.g., sounds of others, distractions, technological problems, etc.). Only three studies to date had completed a study looking at the validity of testing over videoconferencing in the patient's home and this current study adds to this literature, with discrepant results compared to their three studies (Abdolahi et al., 2016; Lindauer et al., 2017; Stillerova et al., 2016).

One could certainly argue that having the participants in their own home was also a great limitation of this study. It meant that the environment could not be controlled for despite prompting the participant to place themselves in a location that would not allow for disruptions, 14.0 percent of individuals in this evaluation experienced some situation that could have impacted their performance on testing (see Table 9). Additionally, as seen below, there were eight individuals that used a device that was not a computer or laptop; these devices likely had smaller screens and could have impacted their engagement with stimuli when the screen was shared. Considering one of the benefits of providing teleneuropsychological services is the extension of services to individuals that live far away from clinics, understanding the limitations of this administration modality is important.

Table 9. Information regarding technology use within teleadministration.

	Group 1 (*N* = 129)
	N (%)
Type of Electronic Device Used	
Desktop Computer	17 (13.2)
Laptop	104 (80.6)
Tablet	3 (2.3)
Other	5 (3.9)
Technical Issues Documented*	
Yes	18 (14.0)
No	111 (86.0)

* = Two individuals lost connection after first task; they were not included in these numbers as they were not included across analyses.

Another large limitation of the study was the confounds placed on data collection by the COVID-19 pandemic to ensure examiner and participant safety. We had four separate groups; one with active data collection over videoconferencing, whereas the other three were comparison groups that were retrospectively collected. Most studies that have looked at the feasibility of providing services through teleneuropsychology have utilized the same group in both administration modalities and they are then able to compare each individual's performance accounting for the setting. Our study did not have the opportunity to do that, thus we were limited to comparing the group means. Despite this, the size of the differences found are interesting and pose additional questions that can be answered in future research.

Furthermore, as this study attempted to look at feasibility of teleneuropsychological assessment, another limitation of the study was the lack of qualitative data collected. While previous literature suggests teleassessment is well-received, no questions were asked related to the participant's experiences and whether they were invested in the study. Gathering this data could have offered additional information as to possible reasonings for the many significant findings across test administration modality.

Finally, there is a limitation in the assessments used in this study as we were limited to previously gathered data for comparison groups. Although this study found nonsignificant findings between administration modality for the administration of the Reading the Mind in the Eyes Test as expected, various concerns have been raised about the limitations of this measure. As such, some of the pictures display the eyes/head at a tilt, providing more information outside of the eye gaze, which is the primary focus of the test. Similarly, the quality of the pictures are not ideal as they appear dated and low quality; some studies have found participants are more likely to select the correct description word when the pictures are brightened (Hallerbäck et al., 2009). Finally, there may be some cultural implications as the images of people in the study are limited to Caucasian individuals and our current study had 38 percent of participants identify as a minority (Kittel, Olderback, & Wilhelm, 2021). These limitations however paint the vast opportunities and necessity for future research within the field of teleneuropsychology.

Future Directions

Future research should look to address the limitations noted above, particularly when evaluating the feasibility of conducting these evaluations with the patients in their homes. One could consider following teleneuropsychology models that create satellite campuses that are nearer to patients, or only offering clinical interviews/feedbacks through videoconferencing. Should assessment want to be offered in the home, there may be an option to use software that allows for monitoring of the patient's environment, similar to the technology used for standardized testing at home (e.g., Graduate Record Examinations; ETS, 2022).

Continued replicability is important as well; while most published studies have shown minimal differences across most cognitive assessments, our current study suggests that research should continue to be conducted to further assess the feasibility of administering neuropsychological measures through video conferencing. Should differences continue to be observed, one could also suggest that future research look to establish normed data for telehealth services, especially if the overall depressed scores we observed in the current study, are consistent.

Conclusions

This study looked to build upon the current literature supporting the utility and feasibility of administrating neuropsychology assessments through videoconferencing. Prior to this study, only cognitive measures had been evaluated and the current research looked to assess whether performance validity testing and social cognition measures would produce similar performances regardless of administration modality. Surprisingly, the results of this study found that numerous cognitive measures did not produce comparable results across testing modalities which suggests continued research is needed to evaluate the effectiveness of administering assessments over videoconferencing. While the majority of performance validity measures did not produce significant differences, the significant findings on Dot Counting Test and Trial 1 of the TOMM illustrates a need for further evaluation and possible inclusion of additional instructions (particularly on DCT). Congruent with expected findings, there were no significant differences on a measure of social cognition; additional research is needed to expand the quantity and quality of measures evaluated. Additionally, the results of this study suggest that continued research and validation of the Raven's Progressive Matrices (shortened form) is recommended as it did not correlate with another measure of generalized intelligence. The findings also suggest the need for further research when testing remotely in an uncontrolled environment and provides an opportunity for further evaluation of performance across cognitive domains through the use of videoconferencing. A major limitation of this study; however, is that the same participants were not used across comparison groups which could account for some of the differences in performance and that the participants were testing from their personal homes, limiting the opportunity for control of distractors. While there were numerous limitations to this study, the results offer a counter to much of the current literature. These results suggest it is imperative that research continue to evaluate the utility of teleneuropsychology and the feasibility of administrating assessments over videoconferencing.

REFERENCES

REFERENCES

Abdelrahman, M. (2021). The indefatigable worker: From factory floor to Zoom avatar. *Critical Sociology, 48*(1), 75-90.

Abdolahi, A., Bull, M. T., Darwin, K. C., Venkataraman, V., Grana, M. J., Dorsey, E. R., & Biglan, K. M. (2016). The feasibility study of conducting the Montreal Cognitive Assessment remotely in individuals with movement disorders. *Health Informatics Journal, 22*(2), 304-311.

Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment, 8*, 323-336.

Alioto, A. G., Kramer, J. H., Borish, S., Neuhaus, J. Saloner, R., Wynn, M., & Foley, J. M. (2017). Long-term test-retest reliability of the California Verbal Learning Test – second edition. *The Clinical Neuropsychologist, 31*(8), 1449-1458.

American Hospital Association (2019, February). *Fact sheet: Telehealth.* https://www.aha.org/system/files/2019-02/fact-sheet-telehealth-2-4-19.pdfhttps://www.aha.org/system/files/2019-02/fact-sheet-telehealth-2-4-19.pdf

American Medical Association. (2020, May). Monthly telehealth regional tracker, May 2020. https://www.fairhealth.org/states-by-the-numbers/telehealth

American Psychiatric Association, & American Telemedicine Association. (2018, April). *Best practices in videoconferencing-based telemental health.* https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/Telepsychiatry/APA-ATA-Best-Practices-in-Videoconferencing-Based-Telemental-Health.pdf

American Psychological Association (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060-1073.

American Psychological Association. (2013). Guidelines for the practice of telepsychology: Joint task force for the development of telepsychology guidelines for psychologists. *The American Psychologist*, *68*(9), 791-800.

American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist, 70*(9), 832-864.

American Psychological Association. (2020, September 2). *Telehealth guidance by state during COVID-19.* APA. https://www.apaservices.org/practice/clinic/covid-19-telehealth-state-summary

American Telemedicine Association. (2018). *Core operational guidelines for telehealth services involving provider-patient interactions.* American Telemedicine Association.

An, K. Y., Kaploun, K., Erdodi, L. A., & Abeare, C. A. (2016). Performance validity in undergraduate research participants: A comparison of failure rates across tests and cutoffs. *The Clinical Neuropsychologist, 31*(1), 193-206.

Anstey, K. J., Dain, S., Andrews, S., & Drobny, J. (2002). Visual abilities in older adults explain age-differences in Stroop and fluid intelligence but not face recognition: Implications for the vison-cognition connection. *Aging, Neuropsychology, and Cognition, 9*(4), 253-265.

APA Practice Organization (2014, Winter). APA approves guidelines for the practice of telepsychology. *Good Practice,* 6-7.

Armistead-Jehle, P., Green, P., Gervais, R., O., & Hungerford, L. D. (2015). An examination of the Word Memory Test as a measure of memory. *Applied Neuropsychology: Adult, 22*(6), 415-426.

Aronson, S. H. (1977). The Lancet on the telephone 1876-1975. *Medical History, 21*(1), 69-87.

Association of State and Provincial Psychology Boards. (2020, September 8). *Temporary/telepsychological practice and COVID-19*. ASPPB. https://cdn.ymaws.com/www.asppb.net/resource/resmgr/covid19/9.8.2020_temporary_&_telepsy.pdf

Axelrod, B., N., & Lamberty, G. J. (2006). The Oral Trail Making Test. In A. M. Poreh (Eds.), *Neuropsychological assessment: A quantified process approach* (45-52). Swets & Zeitlinger.

Axelrod, B. N., & Lamberty, G. J. (2012). The Oral Trail Making Test. In A. M. Poreh (Ed.), *The quantified process approach to neuropsychological assessment* (pp. 45-51). Psychology Press.

Ball, C., & McLaren, P. (1997). The tele-assessment of cognitive state: A review. *Journal of Telemedicine and Telecare, 3*(3), 126-131.

Ball, C., & Puffett, A. (1998). The assessment of cognitive function in the elderly using videoconferencing. *Journal of Telemedicine and Telecare*, *4*(1), 36-38,

Barcellos, L. F., Bellesis, K. H., Shen, L., Shao, X., Chinn, T., Frndak, S., Drake, A., Bakshi, N., Marcus, J., Schaefer, C., & Benedict, R. H. B. (2017). Remote assessment of verbal memory in MS patients using the California Verbal Learning Test. *Multiple Sclerosis Journal, 24*(3), 354-357.

Barhon, L. I., Batchelor, J., Meares, S., Chekaluk, E., & Shores, E. A. (2014). A comparison of the degree of effort involved in the TOMM and the ACS Word Choice Test using a dual-task paradigm. *Applied Neuropsychology: Adult*, *22*(2), 114–123.

Baron-Cohen, S., Bowen, D. C., Holt, R. J., Allison, C., Auyeung, B., Lombardo, M. V., Smith, P., & Lai, M. C. (2015). The "Reading the Mind in the Eyes" test; Complete absence of typical sex difference in ~400 men and women with autism. *PLOS One, 10*(8), e0136521.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test Revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Barry, D., Bates, M. E., & Labouvie, E. (2008). FAS and CFL forms of verbal fluency differ in difficulty: A meta-analytic study. *Applied Neuropsychology, 15*(2), 97-106.

Bastug, G., Ozel-Kizil, E. T., Ssakarya, A., Altintas, O., Kirici, S., & Altunoz, U. (2013). Oral Trail Making task as a discriminative tool for different levels of cognitive impairment and normal aging. *Archives of Clinical Neuropsychology, 28*(5), 411-417.

Batt, K., Shores, E. A., & Chekaluk, E. (2008). The effect of distraction on the Word Memory Test and Test of Memory Malingering performance in patients with a severe brain injury. *Journal of the International Neuropsychological Society, 14*(6), 1074-1080.

Beaman, C. P., Hanczakowski, M., & Jones, D. M. (2014). The effects of distraction on metacognition and metacognition on distraction: Evidence from recognition memory. *Frontiers in Psychology, 5*, 1-13.

Benton, A. L. (1968). Differential behavioural effects in frontal lobe disease. *Neuropsychologia, 6*, 53-60.

Benton, A. L., & Hamsher, K. (1976). *Multilingual Aphasia Examination manual.* University of Iowa.

Bilder, R. M., Postal, K. S., Barisa, M., Aase, D. M., Cullum, C. M., Gillaspy, S. R., Harder, L., Kanter, G., Lanca, M., Lechuga, D. M., Morgan, J. M., Most, R., Puente, A. E., Salinas, C. M., & Woodhouse, J. (2020). InterOrganizational practice committee recommendations/guidance for teleneuropsychology (TeleNP) in response to the COVID pandemic. *The Clinical Neuropsychologist, 0*(0), 1-21.

Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Development of abbreviated nine-item forms of the Raven's Standard Progressive Matrices test. *Assessment, 19*(3), 354-369.

Bird, C. M., Papadopoulou, K., Ricciardelli, P., Rossor, M. N., & Cipolotti, L. (2004). Monitoring cognitive changes: Psychometric properties of six cognitive tests. *British Journal of Clinical Psychology, 43*, 197-210.

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *Clinical Neuropsychologist, 3*(2), 129-136.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessment agreement between two methods of clinical measurement. *Lancet, I,* 307-310.

Bolton, A., & Dorstyn, D. (2015). Telepsychology for posttraumatic stress disorder: A systematic review. *Journal of Telemedicine and Telecare, 21*(5), 254-267.

Bondi, M. W., Serody, A. B., Chan, A. S., Ebserson-Shumate, S. C., Delis, D. C., Hansen, L. A., & Salmon, D. P. (2002). Cognitive and neuropathologic correlates of Stroop Color-Word Test performance in Alzheimer's disease. *Neuropsychology, 16*(3), 335-343.

Boone, K. B., Lu, P., Back, C., King, C., Lee, A., Philpott, L., Shamieh, E., & Warner-Chacon, K. (2002). Sensitivity and specificity of the Rey Dot Counting Test in patients with suspect effort and various clinical samples. *Archives of Clinical Neuropsychology, 17,* 625-642.

Bostantjopoulou, S., Kiosseoglou, G., Katsarou, Z., & Alevriadou, A. (2001). Concurrent validity of the Test of Nonverbal Intelligence in Parkinson's disease patients. *Journal of Psychology, 135*, 205-212.

Bragrad, A., Schelstraete, M. A., Snyers, P., & James, D. G. (2012). Word-finding intervention for children with specific language impairment: A multiple single-case study. *Language, Speech, Hearing, Services, 43,* 222-234.

Brand, J., & McKay, D. (2012). Telehealth approaches to obsessive-compulsive related disorders. *Psychotherapy Research, 22*(3), 306-316.

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.

Britton, B. K., Graesser, A. C., Glynn, S. M., Hamilton, T., and Penland, M. (1983). Use of cognitive capacity in reading: Effects of some context features of text. *Discourse Process, 6,* 39–57.

Buddin, Jr., W. H., Schroeder, R. W., Hargrave, D. D., Von Dran, E. J., Campbell, E. B., Brockman, C. J., Heinrichs, R. J. & Baade, L. E. (2014). An examination of the frequency of invalid forgetting on the Test of Memory Malingering. *The Clinical Neuropsychology, 28*(3), 525-542.

Bujang, M. A., Sa'at, N., Joys, A. R., & Ali, M. M. (2015). *An audit of the statistics and the comparison with the parameter in the population.* American Institute of Physics Conference Proceedings.

Burke, H. R. (1985). Raven's Progressive Matrices (1938): More on norms, reliability, and validity. *Journal of Clinical Psychology, 41*(2), 231-235.

Cahn-Hidalgo, D., Estes, P. W., & Benabou, R. (2020). Validity, reliability, and psychometric properties of a computerized, cognitive assessment test (Cognivue). *World Journal of Psychiatry, 10*(1), 1-11.

Cameron, J. J., & Stinson, D. A. (2019). Gender (mis)measurement: Guidelines for respecting gender diversity in psychological research. *Social and Personality Psychology Compass 13*(11), 1-14.

Carotenuto, A., Rea, R., Traini, E., Ricci, G., Fasanaro, A. M., & Amenta, F. (2018). Cognitive assessment of patients with Alzheimer's disease by telemedicine: Pilot study. *JMIR Mental Health, 5*(2), e31.

Center for Disease Control and Prevention. (2022). *Vital statistics rapid release: Provisional drug overdose death counts.* https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm.

Chafetz, M. D., & Matthews, L. H. (2004). A new interference score for the Stroop test. *Archives of Clinical Neuropsychology, 19*(4), 555-567.

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, 0*(0), 1-9.

Ciemins, E. L., Holloway, B., Coon, P. J., McClosky-Armstrong, T., & Min, S. (2009). Telemedicine and the Mini-Mental State Examination: Assessment from a distance. *Telemedicine and e-Health, 15*, 476-478.

Clement, P. F., Brooks, F. R., Dean, B., & Galaz, A. (2001). A neuropsychology telemedicine clinic. *Military Medicine, 166*(5), 382-384.

Cochrane, H. J., Baker, G. A., & Meudell, P. R. (1998). Simulating a memory impairment: Can amnestic simplicity outperform simulators? *The British Journal of Clinical Psychology, 37*(1), 31-48.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Routledge Academic.

Cohen, M. J., & Stanczak, D. E. (2000). On the reliability, validity, and cognitive structure of the Thurstone Word Fluency Test. *Archives of Clinical Neuropsychology, 15*(3), 267-279.

Craik, F. I. M. (2014). Effects of distraction on memory and cognition: A commentary. *Frontiers in Psychology, 5*, 1-5.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125*(2), 159-180.

Crawford, J. R., Moore, J. W., & Cameron, I. M. (1992). Verbal fluency: A NART-based equation for the estimation of premorbid performance. *British Journal of Clinical Psychology, 31*, 327-239.

Crawford, J. R., Stewart, L. E., Besson, J. A. O., Parker, D. M., & De Lacey, G. (1989). Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extensions. *British Journal of Clinical Psychology, 28*, 267-273.

Cullum, C. M., Gioia, G., & Podell, K. (2020, November). *Taking Neuropsychology Out of the Office: Extending Our Practice through Telehealth Technology* [Conference session]. American Academy of Clinical Neuropsychology Conference 2019, San Diego, California.

Cullum, C. M., 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 20*(10), 1028-1033.

Cullum, C. M., Weiner, M., Gehrmann, H., & Hynan, L. (2006). Feasibility of telecognitive assessment in dementia. *Assessment, 13*(4), 385-390.

Cullum, C., M., Bauer, R., Postal, K., & Taube, D. (2020, April). *Risk Management for Teleneuropsychology* [Continuing Education CourseVideo Webinar]. American Academy of Clinical Neuropsychology.

Davis, J. J., & Wall, J. R. (2014). Examining verbal memory on the Word Memory Test and California Verbal Learning Test – Second Edition. *Archives of Clinical Neuropsychology, 29*, 747-753.

Deary, I. J., Whalley, L. J., & Crawford, J. R. (2004). An “instantaneous” estimate of lifetime’s cognitive change. *Intelligence, 26*, 357-375.

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch’s t-test instead of student’s t-test. *Interntaional Review of Social Psychology, 30*(1), 92-101.

Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch’s *F*-test instead of the Classical *F*-test in one-way ANOVA. *International Review of Social Psychology, 32*(1), 1-12.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test – Second Edition, Adult Version*. San Antonio, Texas: The Psychological Corporation.

Desmond, D. W., Tatemichi, T. K., & Hanzawa, L. (1994). The Telephone Interview for Cognitive Status (TICS): Reliability and validity in a stroke sample. *International Journal of Geriatric Psychiatry, 9*(10), 803-807.

Din, N. C., & Tat Meng, E. C. (2019). Computerized Stroop tests: A review. *Journal of Psychology and Psychotherapy, 9*(1), 1-5.

Englund, C. E., Reeves, D. L., Shingledecker, C. A., Throne, D. R., Wilson, K. P., & Hegge, F. W. (1987). *Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB): Design and specification of the battery* (NHRC Report No. 87-10). Joint Working Group on Drug Dependent Degradation in Military Performance.

ETS. (2022). *At home testing*. https://www.ets.org/gre/revised_general/register/at_home.

Fairchild, R. M., Ferng-Kuo, S. F., Rahmouni, H., & Hardesty, D. (2020). Telehealth increases access to care for children dealing with suicidality, depression, and anxiety in rural emergency departments. *Telemedicine and e-Health.* Advanced online publication.

Farhoudian, A., Radfar, S. R., Ardabili, H. M., Rafei, P., Ebrahimi, M., Zonoozi, A. K., De Jong, C. A. J., … Ekhtiart, H. (2021). A global survey on changes in the supply, price, and use of illicit drugs and alcohol, and related complications during the 2020 COVID-19 pandemic. *Frontiers in Psychiatry, 12,* 1-16.

Fernandes, M. A., & Moscovitch, M. (2000). Divided attention and memory: Evidence of substantial interference effects at retrieval and encoding. *Journal of Experimental Psychology: General, 129*(2), 155-176.

Fernandez-Abascal, E. G., Cabello, R., Fernandez-Berrocal, P., & Baron-Cohen, S. (2013). Test-retest reliability of the ‘Reading the Mine in the Eyes’ test: A one-year follow-up study. *Molecular Autism, 4*(1), 33.

File, T., & Ryan, C. (2013). *Computer and internet use in the United States: 2013*. https://www.census.gov/history/pdf/acs-internet2013.pdf

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,* 189-198.

Franzen, M. D., Tishelman, A. C., Sharp, B. H., & Friedman, A. G. (1987). An investigation of the test-retest reliability of the Stroop color-word test across two intervals. *Archives of Clinical Neuropsychology, 2,* 265-272.

Freeman, J., & Godfrey, H. (2000). The validity of the NART-RSPM index in detecting intellectual delines following traumatic brain injury: A controlled study. *British Journal of Clinical Psychology, 39*, 95-103.

Galusha-Glasscock, J. M., Horton, D. K., Weiner, M. F., & Cullum, C. M. (2015). Video teleconference administration of the Repeatable Battery for the Assessment of Neuropsychological Status. *Arichives of Clinical Neuropsychologist, 31,* 8-11.

Gatz, M., Reynolds, C., Nikolic, J., Lowe, B., Karel, M., & Pedersen, N. (1995). An empirical test of telephone screening to identify potential dementia cases. *International Psychogeriatrics, 7*(3), 429-438.

Glassman, L. H., Mackintosh, M., Talkovsky, A., Wells, S. Y., Walter, K. H., Wickramasinghe, I., & Morland, L. A. (2017). Quality of life following treatment for PTSD: Comparison of videoconferencing and in-person modalities. *Journal of Telemedicine and Telecare.* 1-5.

Glueckauf, R. L., Drude, K. P., Maheu, M. M., Wells, B. A., Wang, Y., Gustafson, D. J., & Nelson, E. (2018). Survey of psychologists' telebehavioral health practices: Technology use, ethical issues, and training needs. *Professional Psychology: Research and Practice, 49*(3), 205-219.

Green, P. (2003). *Green's word memory test for windows: User's manual.* Edmonton, Canada: Green's Publishing.

Greher, M. R., & Wodushek, T. R. (2017a). Performance validity testing in neuropsychology: Scientific basis and clinical application – a brief review. *Journal of Psychiatric Practice, 23*(2), 134-140.

Greher, M. R., & Wodushek, T. R. (2017b). Performance validity testing in neuropsychology: Methods for measurement development and maximizing diagnostic accuracy. *Journal of Psychiatric Practice, 23*(3), 214-220.

Gros, D. F., Veronee, K. Strachan, M., Ruggiero, K. J., & Acierno, R. (2011). Managing suicidality in home-based telehealth. *Journal of Telemedicine and Telecare, 17*(6), 332-335.

Grosch, M. C., Gottlieb, M. C., & Cullum, C. M. (2011). Initial practice recommendations for teleneuropsychology. *The Clinical Neuropsychology, 7,* 1119-1133.

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.

Gunner, J. H., Miele, A. S., Lynch, J. K., and McCaffrey, R. J. (2011). The Albany Consistency Index for the Test of Memory Malingering. *Archives of Clinical Neuropsychology, 27*(1), 1-9.

Hallerbäck, M. U., Lugnegård, T., Hjärthag, F., & Gillberg, C. (2009). The Reading the Mind in the Eyes Test: Test-retest reliability of a Swedish version. *Cognitive Neuropsychiatry, 14*(2), 127-143.

Hammers, D. B., Stolywk, R., Harder, L, & Cullum, C. M. (2020). A survey of international clinical teleneuropsychology service provision prior to and in the context of COVID-19. *The Clinical Neuropsychologist, 0*(0), 1-17.

Hanes, K. R., Andrewes, D. G.>, Smith, D. J., & Pantelis, C. (1996). A brief assessment of executive control dysfunction: Discriminant validity and homogeneity of planning, set shift, and fluency measures. *Archives of Clinical Neuropsychology, 11*, 185-191.

Hanley, J. R., Baker, G. A., & Ledson, S. (1999). Detecting the faking of amnesia: A comparison of the effectiveness of three different techniques for distinguishing simulators from patients with amnesia. *Journal of Clinical Experimental Neuropsychology, 21*, 59-69.

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.

Health Resources & Services Administration. (2019, August). *Telehealth Programs*. https://www.hrsa.gov/rural-health/telehealth

Health Resources & Services Administration. (2019, May). *About FORHP*. https://www.hrsa.gov/rural-health/about-us/index.html

Heaton, R. K. (2003). *Wisconsin Card Sorting test computer version 4.0*. Psychological Assessment Resources.

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance in patients with a traumatic brain injury. *Neuropsychology, 18,* 621-628.

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia, 42*, 1212-1222.

Hidalgo-Mazzei, D., Mateu, A., Reinares, M., Matic, A., Vieta, E., & Colom, F. (2015). Internet-based psychological interventions for bipolar disorder: Review of the present and insights into the future. *Journal of Affective Disorders, 188*(1), 1-13.

Hildebrand, R., Chow, H., Williams, C., Nelson, M., & Wass, P. (2004). Feasibility of neuropsychological testing of older adults via videoconference: Implications for assessing the capacity for independent living. *Journal of Telemedicine and Telecare, 10,* 130-134.

Himle, J. A., Fischer, D. J., Muroff, J. R., Van Etten, M. L., Lockers, L. M., Abelson, J. L., & Hanna, G. L. (2006). Videoconferencing-based cognitive-behavioral therapy for obsessive-compulsive disorder. *Behaviour Research and Therapy, 44*(12), 1821-1829.

Holmqvist, M., Vincent, N., & Walsh, K. (2014). Web- vs. telehealth-based delivery of cognitive behavioral therapy for insomnia: A randomized controlled trail. *Sleep Medicine, 15*(2), 187-195.

Hunt, C., Borgida, E., & Lavine, H. (2012). Social cognition. *Encyclopedia of Human Behavior,* 456-462.

Hutchings, R., Hodges, J. R., Piguet, O., & Kumfor, F. (2015). Why should I care? Dimensions of socio-emotional cognition in younger-onset dementia. *Journal of Alzheimer's Disease, 48*(1), 135-147.

Institute of Medicine, (1996). *Telemedicine: A guide to assessing telecommunications in health care* (M. J. Field, Ed.). National Academies Press.

Jacobson, 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.

Javed, A., & Charles, A. (2018). The importance of social cognition in improving functional outcomes in schizophrenia. *Frontiers in Psychiatry, 9*, 1-14.

Johnson-Selfridge, M. T., Zalewski, C., & Aboudarham, J. F. (1998). The relationship between ethnicity and word fluency. *Archives of Clinical Neuropsychology, 13*, 319-325.

Kaemmerer, T., & Riordan, P. (2016). Oral adaptation of the Trail Making Test: A practical review. *Applied Neuropsychology: Adult*, *23*(5), 384-389.

Kane, R. L., & Kay, G. G. (1992), Computerized assessment in neuropsychology: A review of tests and test batteries. *Neuropsychological Review, 3*(1), 1-117.

Kane, R. L., Short, P., Spies, W., & Flynn, C. F. (2005). Development and validation of the Spaceflight Cognitive Assessment Tool for Windows (WinSCAT). *Aviation, Space, and Environmental Medicine, 76*(6), B183-B191.

Kapur, N. (1994). The coin-in-the-hand test: A new "bed-side" test for the detection of malingering in patients with suspected memory disorder. *Journal of Neurology, Neurosurgery, & Psychiatry, 57*(3), 385-386.

Kay, G. G. (1995). *CogScreen: Aeromedical edition.* Cognitive Research, Inc.

Kent, J., & Plomin, R. (1987). Testing specific cognitive abilities by telephone and mail. *Intelligence, 11*, 391-400.

Keo-Meier, C. L., & Fitzgerald, K. M. (2017). Affirmative psychological testing and neurocognitive assessment with transgender adults. *Psychiatric Clinics of North America, 40*(1), 51-64.

Kim, E. H., Gellis, Z. D., Bradway, C. K., & Kenaley, B. (2018). Depression care services and telehealth technology use for homebound elderly in the United States. *Aging & Mental Health, 23*(9), 1164-1173.

Kirkland, K. T., Peck, D. F., & Bennie, L. (2000). The consistency of neuropsychological assessments performed via telecommunication and face to face. *Journal of Telemedicine and Telecare, 6,* 147-151.

Kittel, A. F. D., Olderbak, S., & Wilhelm, O. Sty in the mind's eye: A meta-analytic investigation of the nomological network and internal consistency of the "Reading the Mind in the Eyes" test. *Assessment, 29*(5), 872-895.

Knight, R. G., McMahon, J., Green, T. J., and Skeaff, C. M. (2010). Regression equations for predicting scores of persons over 65 on the Rey Auditory Verbal Learning Test, the mini-mental state examination, the trail making test, and semantic fluency measures. *British Journal of Clinical Psychology, 45*(3), 393-402.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163.

Kowalczyk, A., McDonald, S., Cranney, J., & McMahon, M. (2001). Cognitive flexibility in the normal elderly and in persons with dementia as measured by the Written and Oral Trail Making Tests. *Brain Impairment, 2*(1), 11-21.

Kreutzer, J. S., DeLuca, J., & Caplan, B. (Eds.). (2011). *Encyclopedia of Clinical Neuropsychology.* Springer-Verlag.

Lacey, M. A., Gore, P. A., Plisken, N. H., Henry, G. K., Heilbronner, R. L., & Hammer, D. P. (1996). Verbal fluence task equivalence. *Clinical Neuropsychologist, 10,* 305-308.

Lathan, C., Spira, J. L., Bleiberg, J., Vice, J., & Tsao, J. W. (2013). Defense Automated Neurobehavioral Assessment (DANA) – psychometric properties of a new field – deployable neurocognitive assessment tool. *Military Medicine, 178*(4), 365-371.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science, 19*(3), 143–148.

Lawes-Wickwar, S., McBain, H., & Mulligan, K. (2018). Application and effectiveness of telehealth to support severe mental illness management: Systematic Review. *JMIR Mental Health*, *5*(4), 1-13.

Leach, L., Kaplan, E., Rewilak, D., Richards, B., & Proulx, B. B. (2000). *Kaplan Baycrest Neurocognitive Assessment.* The Psychological Corporation.

Levine, A. J., Miller, E. N., Becker, J. T., Selnes, O. A., & Cohen, B. A. (2004). Normative data for determining significance of test-retest differences on eight common neuropsychological instruments. *The Clinical Neuropsychologist, 19,* 373-384.

Lezak, M. D., Hoeiwson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological Assessment, Fifth Edition*. Oxford University Press.

Lichstein, K. L., Scogin, F., Thomas, S. J., DiNapoli, E. A., Dillon, H. R., & McFadden, A. (2013). Telehealth cognitive behavior therapy for co-occurring insomnia and depression symptoms in older adults. *Journal of Clinical Psychology, 69*(10), 1056-1065.

Lin, L. A., Casteel, D., Shigekawa, E., Weyrich, M. S., Roby, D. H., & McMenamin, S. B. (2019). Telemedicine-delivered treatment interventions for substance use disorders: A systematic review. *Journal of Substance Abuse Treatment, 101*, 38-49.

Lindauer, A., Seelye, A., Lyons, B., Doge, H. H., Mattek, N., Mincks, K., Kaye, J., & Erten-Lyons, D. (2017). Dementia care comes home: Patient and caregiver assessment via telemedicine. *The Gerontologist*, *57*(5), 1-9.

Loh, P. K., Ramesh, P., Maher, S., Saligari, J., Flicker, L., & Goldswain, P. (2004). Can patients with dementia be assessed at a distance? The use of telehealth and standardised assessments. *Internal Medicine Journal, 34*(5), 239-242.

Loh, P., Donaldson, M., Flicker, L., Maher, S., & Goldswain, P. (2007). Development of a telemedicine protocol for the diagnosis of Alzheimer's disease. *Journal of Telemedicine and Telecare, 13*, 90-94.

Lovell, M. (2007). *ImPACT 2007 (6.0) software user's manual.* ImPACT Applications Inc.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203.

Marra, D. E., Hamlet, K. M., Bauer, R. M., & Bowers, D. (2020a). Validity of teleneuropsychology for older adults in response to COVID-19: A systematic and critical review. *The Clinical Neuropsychologist, 0*(0), 1-42.

Marra, D. E., Hoelzle, J. B., Davis, J. J., & Schwartz, E. B. (2020b). Initial changes in neuropsychologists clinical practice during the COVID-19 pandemic: A survey study. *The Clinical Neuropsychologist, 0*(0), 1-16.

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professions. *The Clinical Neuropsychologist, 29*(6), 741-776.

Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N., & Okut, H. (2020). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist, 34*(1), 88-119.

Massoudi, B., Holvast, F., Bockting, C. L., Burger, H., & Blanker, M. H. (2019). The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: A systematic review and meta-analysis. *Journal of Affective Disorders, 245*, 728-743.

Matthews, D. J. (1988). Raven's Matrices in the identification of giftedness. *Roeper Review, 10*(3), 159-162.

May, C. P., & Hasher, L. (1998). Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 363-379.

McArdle, B. (2021). *Teleneuropsychology in the COVID-19 pandemic: Developing and testing the convergent validity of the eSPANS* (Unpublished doctoral dissertation). University of Leicester.

McCaffrey, R. J., O'Bryant, S. E., Ashendorf, L., & Fisher, J. M. (2003). Correlations among the TOMM, Rey-15, and MMPI-2 validity scales in a sample of TBI litigants. *Journal of Forensic Neuropsychology*, *3*(3), 45-53.

McCallum, S., Bracken, B., & Wasserman, J. (2000). *Essentials of non-verbal assessment.* John Wiley & Sons.

McCaul, C., Boone, K. B., Ermshar, A., Cottingham, M., Victor, T. L., Ziegler, E., Zeller, M. A., & Wright, M. (2018). Cross-validation of the Dot Counting Test in a large sample of credible and non-credible patients referred for neuropsychological testing. *The Clinical Neuropsychologist, 32*(6), 1054-1067.

Menon, A. S., Kondapavalru, P., Krishna, P., Chrismer, J. B., Raskin, A., Hebel, J. R., & Ruskin, P. E. (2001). Evaluation of a portable low cost videophone system in the assessment of depressive symptoms and cognitive function in elderly medically ill veterans. *The Journal of Nervous and Mental Disease, 189*(6), 399-401.

Miller, J. B., Axelrod, B. N., Rapport, L. J., Hanks, R. A., Basehm, J. R., & Schutte, C. (2012). Substitution of California Verbal Learning Test, Second Edition for Verbal Paired Associates on the Wechsler Memory Scale, Fourth Edition. *The Clinical Neuropsychologist, 26*(4), 599-608.

Mills, C. J., Ablard, K. E., & Brody, L. E. (1993). The Raven's Progressive Matrices: Its usefulness for identifying gifted/talented students. *Roeper Review, 15*, 183-186.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions of Psychological Sciences, 21*, 8-14.

Montani, C., Billaud, N., Tyrrell, J., Fluchaire, I., Malterre, C., Lauvernay, N., Couturier, P., & Franco, A. (1997). Psychological impact of a remote psychometric consultation with hospitalized elderly people. *Journal of Telemedicine and Telecare, 3*(3), 140-145.

Mrazik, M., Millis, S., & Drane, D. L. (2010). The Oral Trail Making Test: Effects of age and concurrent validity. *Archives of Clinical Neuropsychology, 25*(3), 236-243.

Multi-Health Systems. (2014). *Conners CPT 3 Use Manual*. Multi-Health Systems Publishing.

Myers, K., & Turvey, C. L. (Eds.). (2013). *Elsevier insights. Telemental health: Clinical, technical, and administrative foundations for evidence-based practice.* Elsevier Publishing.

Nelson, N. W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships between eight measures of suspect effort. *The Clinical Neuropsychologist, 17*, 263-272.

Nesselroade, J. R., Pedersen, N. L., McClearn, G. E., Plomin, R., & Bergeman, C. S. (1998). Factorial and criterion validities of telephone-assessed cognitive ability measures: Age and gender comparisons in adult twins. *Research on Aging, 10*(2), 220-234.

Newcombe, F. (1969). *Missile Wounds of the Brain. A Study of Psychological Deficits.* Oxford University Press.

Okada, K. (2013). Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika, 40*, 129-147.

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: Toward a brief form for research and applied settings. *Frontiers in Psychology, 6*, 1-14.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8*(4), 434-447.

Oliveira-Souza, R. D., Moll, J., Passman, L. J., Cunha, F. C.>, Paes, F., Adriano, M. V., Ignacio, F. A., & Marrocos, R. P. (2000). Trail making and cognitive set-shifting. *Arquivos de Neuro-Psiquiatria, 58*(3B), 826-829.

Parks, A. C., Davis, J., Spresser, C. D., Stroescu, I., & Ecklund-Johnson, E. (2021). Validity of in-home teleneuropsychological testing in the wake of COVID-19. *Archives of Clinical Neuropsychology, 36*(6), 887-896.

Parikh, M., Grosch, M. C., Graham, L. L., Hynan, L. S., Weiner, M., Shore, J. H., & Cullum, C. M. (2013). Consumer acceptability of brief videoconference-based neuropsychological assessment in older individuals with and without cognitive impairment. *Clinical Neuropsychologist, 27*(5), 808-817.

Patel, K., Robertson, E., Kwong, A. S. F., Griffith, G. J., Willan, K., Green, M. J., Di Gessa, G., … Katikireddi, S. V. (2022). Psychological distress before and during the COVID-19 pandemic among adults in the United Kingdom base on coordinated analyses of 11 longitudinal studies. *JAMA Network Open, 5*(4), 1-15.

Perednia, D. A., & Allen, A. (1995). Telemedicine technology and clinical applications. *JAMA: The Journal of the American Medical Association, 273*(6), 483-488.

Perez, W. A., Masline, P. J., Ramsey, E. G., & Urban, K. E. (1987). *Unified Tri-services Cognitive Performance Assessment Battery: Review and methodology* (AAMRL Publication No. TR-87-007). Joint Working Group on Drug Dependent Degradation in Military Performance.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ, 316*(7139). 1236-1238.

Peterson, E., & Miller, S. F. (2012). The Eyes Test as a measure of individual differences: How much of the variance reflects verbal IQ? *Frontiers in Psychology, 3*, 1-6.

Pierce, B. S., Perrin, P. B., & McDonald, S. D. (2020). Demographic, organizational, and clinical practice predictors of U.S. psychologists' use of telepsychology. *Professional Psychology: Research and Practice, 51*, 184-193.

Plassman, B. L., Newman, T. T., Welsh, K. A., Helms, M., & Breitner, J. C. S. (1994). Properties of the telephone interview for cognitive status: Application in epidemiological and longitudinal studies. *Neuropsychiatry, Neuropsychology and Behavioral Neurology, 7*(3), 235-241.

Powell, M. R., Gfeller, J. D., Hendricks, B. L., & Sharland, M. (2004). Detecting symptom- and test-coached simulators with the Test of Memory Malingering. *Archives of Clinical Neuropsychology, 19*, 693-702.

Proctor, S. P., Heaton, K. J., Lieberman, H. R., Smith, C. D., Edens, E. N., Kelley, A., Balkin, T. J., Capaldi, V., Doty, T. J., & Quartana, P. J. (2017). *Military cognitive performance and readiness assessment initiative*. https://apps.dtic.mil/dtic/tr/fulltext/u2/1053085.pdf

PSYPACT. (2022, July 19). *PSYPACT Map/States.* PSYPACT. https://psypact.org/page/psypactmap.

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology, 31*(3), 206-230.

Rabin, L. A., Spadaccini, A. T., Brodale, D. L., Grant, K. S., Elbulok-Charcape, M. M., & Barr, W. B. (2014). Utilization rates of computerized tests and test batteries among clinical neuropsychologists in the United States and Canada. *Professional Psychology: Research and Practice, 45*(5), 368-377.

Rae, P. J. L., & Perfect, T. J. (2014). Visual distraction during word-list retrieval does not consistently disrupt memory. *Frontiers in Psychology, 5*, 1-11.

Raven, J. C. (1938). *Progressive Matrices: A perceptual test of intelligence*. Oxford Psychologists Press Ltd.

Raven, J., Raven. J. C. & Court, H. H. (1998/2000). *Raven manual*. Oxford Psychological Press Ltd.

Rees, L., M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment, 10*, 10-20.

Reeves, D. L., Thorne, D. R., Winter, S. L. & Hegge, F. W. (1989). *The Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB): Hardware/software design and specifications* (NAMRL Publication No. SR89-1). Joint Working Group on Drug Dependent Degradation in Military Performance.

Reeves, D. L., Winter, K. P., Bleibery, J., & Kane, R. L. (2007). ANAM genogram: Historical perspectives, description, and current endeavors. *Archives of Clinical Neuropsychology, Supp 1*, S15-37.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8*(3), 271-276.

Reitan, R. M. & Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological test Battery: Theory and clinical interpretation.* Neuropsychology Press.

Resch, J., Driscoll, A., McCaffrey, N., Brown, C., Ferrara, M. S., Macciocchi, S., Baumgartner, T., & Walpert, K. (2013). ImPACT test-retest reliability: Reliably unreliable? *Journal of Athletic Training, 48*(4), 506-511.

Ricker, J. H., & Axelrod, B. N. (1994). Analysis of an oral paradigm for the Trail Making Test. *Assessment, 1*(1), 47-52.

Riedl, R. (2021). On the stress potential of videoconferencing: Definition and root causes of Zoom fatigue. *Electronic Markets, 32*, 153-177.

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science, 29*(8), 1358-1369.

Roca, M., Manes, F., Chade, A., Gleichgerrcht, E., Gershanik, O., Arevalo, G. G., Torralva, T., & Duncan, J. (2012). The relationship between executive functions and fluid intelligence in Parkinson's disease. *Psychological Medicine, 42*(11) 2445-2452.

Roccaforte, W. H., Burke, W. J., Bayer, B. L., & Wengel, S. P. (1992). Validation of a telephone version of the Mini-Mental State Examination. *Journal of the American Geriatrics Society, 40*(7), 697-702.

Ross, T. P. (2003). The reliability of cluster and switch scores for the COWAT. *Archives of Clinical Psychology, 18*, 153-164.

Ross, T. P., Calhoun, E., Cox, T., Wenner, C., Kono, W., & Pleasant, M. (2007). The reliability and validity of qualitative scores for the Controlled Oral Word Association Test. *Archives of Clinical Neuropsychology, 22*, 475-488.

Ross, T. P., Foard, E. F., Hiott, F. B., & Vincent, A. L. (2003). The reliability of production strategy scores for the Ruff Figural Fluency Test. *Archives of Clinical Neuropsychology, 18*, 879-891.

Ruchinskas, R. (2003). Limitations of the Oral Trail Making Test in a mixed sample of older individuals. *The Clinical Neuropsychologist, 17*(2), 137-142.

Sammons, M. T., VandenBos, G. R., & Martin, J. N. (2020). Psychological practice and the COVID-19 crisis: A rapid response survey. *Journal of Health Service Psychology, 46*, 51-57.

Sauzeon, H., Raboutet, C., Rodrigues, J., Langevin, A., Schelstraete, M. A., Feyereisen, P., Hupet, M., & N'Kaoua, B. (2011). Verbal knowledge as a compensation determinant of adult age differences in verbal fluency tasks over time. *Journal of Adult Development*, *18*, 144-154.

Scarpina, F., & Tagini, S. (2017). The Stroop color and word test. *Frontiers in Psychology*, *8*, 1-8.

Scheiter, K., Gerjets, P., & Heise, E. (2014). Distraction during learning with hypermedia: Difficult tasks help to keep task goals on track. *Frontiers in Psychology, 5*, 1-12.

Schlegel, R. E., Gilliland, K., & Crabtree, M. S. (1992). *Development of the UTC-PAB normative database* (Technical Report ADA271319). University of Oklahoma.

Schretlen, D. (1997). *Brief Test of Attention professional manual*. Psychological Assessment Resources.

Schroeder, R. W., & Martin, P. K. (2021). *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance*. Guilford Press.

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. *Assessment, 19*(1), 21-30.

Schwartz, E. S., Erdodi, L., Rodrigues, N., Ghosh, J. J., Curtain, J. R., Flashman, L. A., & Roth, R. M. (2016). CVLT-II forced choice recognition trial as an embedded validity indicator: A systematic review of the evidence. *Journal of the International Neuropsychological Society*, *22*, 851-858.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*(772), 1-10.

Shingleton, R. M., Richards, L. K., & Thompson-Brenner, H. (2013). Using technology within the treatment of eating disorders: A clinical practice review. *Psychotherapy*, *50*(4), 576-582.

Shura, R. D., Taber, K. H., Armistead-Jehle, P., Denning, J. H., & Rowland, J. A. (2019). Effects of distraction on performance validity: A pilot study with veterans. *Archives of Clinical Neuropsychology*, *34*(8), 1432-1437.

Sloan, D. M., Gallaghet, M. W., Feinstein, B. A., Lee, D. J., & Pruneau, G. M. (2011). Efficacy of telehealth treatments for posttraumatic stress-related symptoms: A meta-analysis. *Cognitive Behaviour Therapy, 40*(2), 111-125.

Social intelligence test. (2022). Lab in the Wild. Retrieved May 25, 2022, from http://socialintelligence.labinthewild.org/mite/.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35*(5), 401-426.

Stolwyk, R., & Hammers, D. B. (2020, February 5). *A Practical Guide to the Ethics and Practice of Rural Health and Teleneuropsychology* [Conference session]. International Neuropsychological Society Conference 2020, Denver, Colorado.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary, third edition.* Oxford University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643-662.

Sweet J. J., Meyer D. G., Nelson N. W., Moberg P. J. (2011). The TCN/AACN 2010 “Salary Survey”: Professional practices, beliefs, and incomes of U.S. neuropsychologists. *The Clinical Neuropsychologist, 25*, 12–61.

Takacs, A., Kobor, A., Tarnok, Z., & Csepe, V. (2013). Verbal fluency in children with ADHD: Strategy using and temporal properties. *Child Neuropsychology, 20*(4), 415-419.

Teichner, G., & Wagner, M. T. (2004). The Test of Memory Malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology*, *19*, 455-464.

The Psychological Corporation. (1997/2002). *WAIS-III / WMS-III Technical Manual.* The Psychological Corporation.

Thurstone, L. L. (1938). *Primary Mental Abilities*. University of Chicago Press.

Tombaugh, T. (1996). *Test of Memory Malingering (TOMM).* Multi Health Systems.

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluence: FAS and animal naming. *Archives of Clinical Neuropsychology*, *14*, 167-177.

Trittschuh, E. H., Parmenter, B. A., Clausell, E. R., Mariano, M. J., & Reger, M. A. (2018). Conducting neuropsychological assessment with transgender individuals. *The Clinical Neuropsychologist*, 1-18.

Tröster, A. I., Paolo, A. M., Glatt, S. L., Hubble, J. P., & Koller, W. C. (1995). “Interactive video conferencing” in the provision of neuropsychological services to rural areas. *Journal of Community Psychology, 23,* 85-88.

United Nations. (2022). *Everyone included: Social impact of COVID-19.* https://www.un.org/development/desa/dspd/everyone-included-covid-19.html#:~:text=If%20not%20properly%20addressed%20through,the%20medium%20and%20long%20term.

U.S. Department of Commerce. (2017). *Census bureau, 2017 American community survey (ACS) public use microdata sample (PUMS) data.* https://nces.ed.gov/programs/digest/d18/tables/dt18_702.60.asp

Uttl, B. (2002). North American Adult Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology*, *24*(8), 1123-1137.

Uttl, B B., & Graf, P. (1997). Color Word Stroop test performance across the life span. *Journal of Clinical and Experimental Neuropsychology*, *19*, 405-420.

Vahia, I. V., Ng, B., Camacho, A., Cardenas, V. Cherner, M., Depp, C. A., Palmer, B. W., Jeste, D. V., & Agha, Z. (2015). Telepsychiatry for neurocognitive testing in older rural Latino adults. *The American Journal of Geriatric Psychiatry*, *23*(7), 666-670.

Vanden Bos, G. R., & Williams, S. (2000). The internet versus the telephone: What is telehealth anyway? *Professional Psychology: Research and Practice*, *31*(5), 490-492.

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The "Reading the Mind in the Eyes" test: A systematic review of psychometric properties and a validation study in Italy.

Vestal, L., Smith-Linde, L., Hicks, G., Hutton, T., & Hart Jr., J. (2006). Efficacy of language assessment in Alzheimer's disease: Comparing in-person examination and telemedicine. *Clinical Interventions in Aging*, *1*(4), 467-471.

Wadsworth, H. E., Dhima, K., Womack, K. B., Hart, J., Weiner, M. F., Hynan, L. S., & Cullum, C. M. (2018). Validity of teleneuropsychological assessment in older patients with cognitive disorders. *Archives of Clinical Neuropsychology*, *33*(8), 1040-1045.

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*, *31*(5), 420-425.

Wais, P. E., & Gazzaley, A. (2014). Distractibility during retrieval of long-term memory: Domain-general interference, neural networks, and increased susceptibility in normal aging. *Frontiers in Psychology*, *5*, 1-12.

Waterfall, R. C. (1979). Automating standard intelligence tests. *Journal of Audiovisual Media in Medicine*, *2*, 21-24.

Webb, A., Holmes, J. E., & Peta, J. L. (2016, April). Assessment norms for gender and implications for transgender, nonbinary populations. *Division 44 Newsletter*. https://www.apadivisions.org/division-44/publications/newsletters/division/2016/04/nonbinary-populations.

Welch, B. L. (1947). The generalization of student's problem when several different population variances are involved. *Biometrika*, *34*(1/2), 28-35.

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso M. R., & Roper, B. (2015). Verbal fluence: Language or executive function measure? *Applied Neuropsychology: Adult, 23*(1), 29-34.

Wilde, N., & Strauss, E. (2002). Functional equivalence of WAIS-III/WMS-III Digit and Spatial Span under forward and backward recall conditions. *The Clinical Neuropsychologist, 16*, 322-330.

Wittson, C. L., & Affleck, D. C., & Johnson, V. (1961). Two-way television in group therapy. *Psychiatric Services, 12*(11), 22-23.

Wittson, C. L., & Benschoter, R. (1972). Two-way television: Helping the medical center reach out. *American Journal of Psychiatry, 129*(5), 624-627.

Woods, S. P., Delis, D. C., Scott, J. C., Kramer, J. H., & Holdnack, J. A. (2006). The California Verbal Learning Test – Second Edition: Test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms. *Archives of Clinical Neuropsychology, 21*, 413-420.

World Health Organization. (2022, March 2). *COVID-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide: Wake-up call to all countries to step up mental health services and support.* Retrieved from https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide.

Wrape, E. R., & McGinn, M. M. (2018). Clinical and ethical considerations for delivering couple and family therapy via telehealth. *Journal of Marital and Family Therapy, 45*(2), 296-308.

Wright, C. V., Beattie, S. G., Galper, D. I., Church, A. S., Bufka, L. F., Brabender, V. M., & Smith, B. L. (2017). Assessment practices of professional psychologists: Results of a national survey. *Professional Psychology: Research and Practice, 48*(2), 73–78.

Yiğit, S., & Mendes, M. (2018). Which effect size measure is appropriate for one-way and two-way ANOVA models? A Monte Carlo simulation study. *REVSTAT – Statistical Journal, 16*(3), 295-313.

Young, L. B. (2011). Telemedicine interventions for substance-use disorder: A literature review. *Journal of Telemedicine and Telecare, 18*(1), 47-53.

Zoom Video Communications, Inc. (2019). *Privacy & security at Zoom Video Communications. Zoom Video Communications, Inc.* https://zoom.us/docs/en-us/privacy-and-security.html.