The Effects of Mask-Wearing and Teleneuropsychological Assessment on People's Memory Performance on Neuropsychological Tests

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Master of Science in Psychology

at

The University of Waikato

by

Tiffany Rich

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2023

Abstract

In the wake of the Coronavirus disease (COVID-19), the standardised conditions of neuropsychological assessments were greatly impacted to prevent the spread of COVID-19. Neuropsychological assessment conditions were altered to be conducted remotely on Zoom (teleneuropsychological) or continued in-person with the use of face masks. The changes to standardised assessment conditions led to an increase in people's worries around assessment, made it harder to hear the assessor which led people to increase their listening effort and created more ambiguous non-verbal social cues (Bottalico et al., 2020; Aguilar & Leguizamon, 2021). These additional assessment aspects may be leading to increased pressure on people's cognitive load by taking up more of their cognitive resources and attention, then leaving less cognitive resources to focus on the assessment. Moreover, memory tests require high rates of attention and focus to accurately encode, manipulate and retrieve information that is presented. However, the added pressure on cognitive load leaves less available cognitive resources to focus on the memory tests, which in turn may influence people's performance on these tests. When cognitive load is overloaded, people have issues with attention, encoding and later retrieving or remembering that same information (Lezak et al., 2004). To determine whether the changes to assessment conditions (test environment and mask-wearing) impact people's memory performance, we conducted a 2x2 mixed methods design (between-subjects factor: test environment (zoom/in-person), within-subjects factor: mask-wearing (mask/no mask). Our sample consisted of 64 participants, aged 18-59, with 43 females, 20 males and 1 participant who preferred to self-describe. Participants were included based on a screening criteria and were required to complete a range of neuropsychological tests for two sessions, either via Zoom or in-person, 1-2 weeks apart. All participants completed one session with the assessor wearing a mask and one session without. Participants scores were compared on the Verbal Learning, Digit Span and Verbal Fluency subtests between the test environments (Zoom/In-person) and mask-wearing (Mask/no mask) variables, to assess their impacts on people's memory functioning. Results showed no significant differences between participants' performance on these tests between the in-person and teleneuropsychological assessments or mask-wearing compared to no mask. A statistically significant but not clinically significant interaction effect was found on Verbal Learning trial 5, where participants performed the worst in the Zoom condition when the assessor was wearing a mask. This result was not clinically significant and overall, our results suggest teleneuropsychological assessment and mask wearing during assessment do not impact people's memory performance and, therefore, are valid and reliable methods of assessment in a population aged 18-59 years with no cognitive impairments. However, we did find that the alternate forms used for the Verbal Learning subtest were not equivalent. Participants' performed statistically significantly worse and recalled fewer words on the Crawford Verbal Learning test compared to the Rey Auditory Verbal Learning Test (RAVLT). This result was not clinically significant but is a clinically meaningful finding for neuropsychologists in New Zealand to consider when administering alternate forms of the RAVLT test.

Acknowledgements

I would like to express my deepest appreciation and gratitude to my supervisor Nicola Starkey for her continuous support, invaluable knowledge and advice. A special thank you to my mum, partner, family and friends for their encouragement and belief in me throughout this journey. I am also grateful for my friends who were on this journey with me, for their laughter and moral support during the stressful times. Lastly, I would also like to thank the study participants for their time and participation, the study could not have been completed without it.

Table of Contents

List of Figures

List of Tables

Introduction & Literature Review

Neuropsychology is a branch of psychology that is concerned with the connection between the brain and behaviour. Specifically, Clinical Neuropsychology is a field of neuropsychology that specialises in the assessment and diagnosis of brain impairments or disorders, and studies the behavioural expression of those impairments (Groth-Marnat & Wright, 2016). Neuropsychological assessments are typically conducted when someone has an injury or disorder that affects their brain, such as, a traumatic brain injury, medical or genetic disorder, neurological concerns or other disorders, including, movement disorders, dyslexia, ADHD, autism or psychiatric disorders, anxiety or depression (Semrud-Clikeman & Swaiman, 2017). The purpose of neuropsychological assessments is to assess the person's impairment or injury in the hope of providing information on the impact of their condition on their cognitive daily functioning and how that may affect other areas of their functioning (Groth-Marnat & Wright, 2016). Assessments may be used for diagnosis and identification of possible neurological disorders and to provide information on the person's cognitive strengths and weaknesses to help design interventions for the individual and aid in management, living with their impairment (Laatsch, 2002; Groth-Marnat & Wright, 2016). These interventions include designing programs, evaluating the effectiveness of drug treatments, physical training and other therapy (Cohen & Swerdlik, 2005). However, the valuable information gathered from neuropsychological tests can also be used for information on legal matters and for research (Groth-Marnat & Wright, 2016).

There are a range of different categories of neuropsychological tests that assess different areas of brain functioning. There are tests specified to look at the functioning of a person's memory, intelligence, language, executive function, visuospatial, dementia-specific and multiple functions (Groth-Marnat & Wright, 2016). The specific category of tests administered are selected based on the reasons for the assessment and the referral question. Referrals can be made by a range of different people — physicians, Accident Compensation Corporation (ACC), General Practitioners (GPs), self-referrals, court — who are concerned about the person's cognitive functioning and will outline why the assessment needs to be done.

In New Zealand, neuropsychological assessments are important to provide accurate results for diagnosis but also for the ability to claim funding from ACC in cases where people have suffered from a brain injury (Accident Compensation Corporation, n.d). Usually, neuropsychological assessments include a series of tests undertaken in standardised conditions, where the client and assessor are in the same room, face-to-face, usually at a clinic. This provides the assessor with a better opportunity to build rapport with the client. Rapport is defined as the relationship between the assessor and client and is critical for accurate assessment (Cohen & Swerdilk, 2005). Rapport helps the client feel comfortable and at ease during the assessment, which helps to avoid the interference of anxiety or nerves on performance. Additionally, distractions in the environment such as excessive noise, temperature (too hot, too cold), interruptions, glaring sunlight, other people in the room, inadequate ventilation should be avoided (Cohen & Swerdilk, 2005). These conditions are important to provide a valid and reliable neuropsychological assessment, as it gives the client the ability to perform at their optimal performance. Valid and reliable assessments are important to accurately determine the impact of the person's injury or impairment on their cognitive and behavioural functioning and help to plan effective interventions for support. Although not all variables that impact the clients performance can be controlled; test anxiety, willingness to cooperate and comprehend test instructions, hunger/fullness, emotional distress, physical pain, alertness, importance of portraying image of themselves, prior coaching and predisposed to agreeing or disagreeing to stimulus statements (Cohen & Swerdilk, 2005).

In the wake of the Coronavirus disease (COVID-19), neuropsychological assessments could not be administered under the recommended assessment conditions to limit the spread of COVID-19. Instead, the adoption of telehealth greatly increased, and the tests began to be administered remotely, on video conferencing software, for example, on zoom (Parks et al., 2021). This remote, virtual, administration method is often referred to as teleneuropsychology. During teleneuropsychological assessments, the client may complete the assessment from the comfort of their own home or residence or in the clinic but in a separate room from the assessor. Additionally, some neuropsychological tests were adapted into online versions and clients were sent any materials they needed for the assessments in the mail (Tailby et al., 2020). As COVID-19 restrictions began to ease, the tests continued being administered in-person, however, the assessor was required to wear personal protective equipment (PPE) including masks, coupled with the client wearing a mask too, for their safety. While COVID-19 is still around, mask wearing is still mandatory in healthcare settings in New Zealand. These new ways of administering neuropsychological assessments have changed the standardised test conditions that have been validated and provide reliable information. What we do not know is the impact these new assessment conditions have on people's performance on neuropsychological tests and if they still provide accurate and reliable assessment results.

Systematic Review of the Literature on Teleneuropsychology

To help understand what is already reported in the literature about teleneuropsychological assessments and the impact of face mask usage on people's test performance, we will first look at the literature on teleneuropsychology as a method of neuropsychological assessment for older adults. A systematic review article by Marra et al. (2020) concluded that teleneuropsychological assessment is a valid method of testing. Marra et al. (2020) collected relevant articles from 2016 who recruited older adults to undergo face-to-face (in-person) and teleneuropsychology assessments in a counterbalanced design. The adults in these studies ranged from healthy controls to people with memory disorders, movement disorders, stroke patients, psychiatric patients and mixed clinical groups (Marra et al., 2020). Looking at a range of different neuropsychological tests such as Digit Span Forwards and Backwards, Letter and Category Fluency and Verbal Learning, Marra et al. (2020) found the majority had no differences between using them in face-to-face methods compared to online teleneuropsychology methods. Therefore, concluding teleneuropsychology is a reliable method to assess cognitive functioning in older adults. A second review article comparing teleneuropsychological assessment to face-to-face measures by Sharma et al. (2021) also concluded that it has been a successful way of conducting assessment during the COVID-19 pandemic. Teleneuropsychological assessment in comparison to in-person is valid for Digit Span, Verbal Fluency and List Learning measures (Brearly et al., 2017 as cited in Sharma et al., 2021). Moreover, the same conclusion was reached for common screening measures, such as, the Montreal Cognitive Assessment (MoCA) (Chapman et al., 2019 as cited in Sharma et al., 2021). Teleneuropsychological assessment was also suggested to be a valid assessment and evaluation for people who have suffered from a stroke, cancer, dementia, mild cognitive impairment, major depressive disorder, anxiety, post-traumatic stress disorder, autism and for school-based cognitive and achievement testing (Galusha-Glasscock et al., 2016; Godleski et al., 2012; Grosch et al., 2011; Hess & Audebert, 2013; Parmanto et al., 2013; Smith et al., 2017; Wadsworth et al., 2017 as cited in Sharma et al., 2021). The satisfaction and acceptance ratings were also found to be high for 98% of clients, some preferring it over face-to-face assessments (Dang et al., 2018; DeYoung & Shenal, 2019; Feenstra et al., 2018; Harrell et al., 2014; Parikh et al., 2013 as cited in Sharma et al., 2021).

However, we do know that there may be barriers to teleneuropsychological testing as 261 licensed clinical neuropsychologists in the United States of America (US) reported in an online survey that they had little knowledge of teleneuropsychology, accessibility and software (Rochette et al., 2021). Additional barriers reported were personal/family stressors, lack of experience, ethical concerns, lack of private office space at home, lack of virtual materials and loss of people to assist with testing (Rochette et al., 2021). Therefore, the majority of the clinical neuropsychologists were unsatisfied or had no preference of using telelneuropsychology as a method of assessment compared to in-person (Rochette et al., 2021). These results suggest that teleneuropsychology may not always be a feasible method.

Although a few review articles found no difference between neuropsychological assessment methods of face-to-face (in-person) and teleneuropsychology, claiming it as a valid and reliable method of testing, this could be due to the type of neuropsychological tests used in those studies. Additionally, because the above results are from review articles, they look at the literature as a whole, and because the findings in the studies are not always consistent and they use varied populations and tests, the review summary can overlook some important information. Looking specifically at the Verbal Learning test, Marra et al. (2020) reported that one study found a significant difference in participants' scores between face-to-face (in-person) and teleneuropsychological methods, but because three other articles found no difference, overall, Verbal Learning was said to have moderate validity and no difference between testing methods. Moreover, a meta-analysis and systematic review done by Brearly et al. (2017) looked at the impact of teleneuropsychological administration on neuropsychological test performance. They found that the literature looked at a diverse range of participants including participants with no impairments, participants being treated in a medical unit, participants in inpatient and outpatient psychiatric or substance use treatment, participants with mild cognitive impairment, Alzheimer's disease or other dementias (Brearly et al., 2017). Overall, there were no significant difference in test scores between teleneuropsychological and in-person administration methods for these participants (Brearly et al., 2017). However, there were diverse results when looking closely at the studies individually, as some studies did find that there were a significant change in scores when using teleneuropsychological methods, both higher and lower scores were found on teleneuropsychological assessment compared to in-person (Svenn et al., 2003; Kirkwood et al., 2000; Montani et al., 1997). The difference in scores suggests that the test environment may alter people's test performance for people with no impairments, people with history of drug and alcohol abuse and hospitalised elderly people (Svenn et al., 2003; Kirkwood et al., 2000; Montani et al., 1997). The tests compared between administration methods in the studies were the Mini-Mental State Examination, Clock Drawing Test, Logical Memory, Seashore Rhythm Test, National Adult Reading Test (NART) and sections of the Adult Memory and Information Processing Battery (AMIPB). Moreover, some neuropsychological tests may also be harder to convert to virtual materials and virtual testing and therefore, people's performance on these tests may be more heavily affected by the change. Additionally, some tests that include visual or auditory components may be harder to administer and complete online due to additional factors such as audio quality, visual stimulus quality and internet connection stability, impacting people's performance.

This next section reviews findings from studies looking at the effects of teleneuropsychological assessment on neuropsychological test performance. I will be focusing on three key cognitive domains assessed by neuropsychologists, intelligence, executive functioning and memory. A summary of research findings in the literature is presented in Table 1.

Effects of Teleneuropsychological Assessment on Specific Cognitive Domains

WAIS-IV

The majority of the previous literature around the validity of teleneuropsychological methods of testing compared to in-person, focus on intelligence tests, commonly from the Weshler Adult Intelligence Scale 4th edition (WAIS-IV). The WAIS-IV is a test battery that contains tests that measure intelligence and cognitive ability. These tests measure aspects of intelligence, such as, intellectual functioning, incorporating verbal, analogical and sequential and quantitative reasoning, also, working memory and psychomotor processing speed (Clime & Rostad, 2011). The section below will cover subtests that measure visuospatial and abstract reasoning and world knowledge and concept formation and then will move on to the teleneuropsychology literature for other cognitive domains (executive functioning and memory).

Visuospatial and Abstract Reasoning

The studies in the literature compare participants' scores on the individual neuropsychological subtests between administration methods. Firstly, literature will be reviewed on the Matrix Reasoning subtest from the WAIS-IV that measures visuospatial and abstract reasoning (Groth-Marnat & Wright, 2016). Previous research found that participants with and without cognitive impairments did not have significantly different performance on this subtest when comparing their scores across in-person and teleneuropsychological methods of testing (Marra et al., 2020; Aguilar & Leguizamon, 2021; Mahon et al., 2021).

Next, the Similarities subtest from the WAIS-IV measures abstract reasoning and verbal concept formation (Groth-Marnat & Wright, 2016). However, it also has a component of long-term memory as people have to think abstractly about knowledge learned and express refined responses (Groth-Marnat & Wright, 2016). There has been little research on the comparison between virtual and face-to-face methods of testing for the Similarities subtest. However, a study by Mahon et al. (2021), found that there were no significant differences on participants' performance on this test between the methods of assessment.

Moving on to the Block Design subtest from the WAIS-IV, it is a subtest that measures non-verbal problem solving skills and visuospatial functioning (Groth-Marnat & Wright, 2016). This subtest requires the assessor to re-create a pattern shown on a page, using blocks and for the participant to create that pattern using the blocks provided in front of them. Additionally, the first few items require the test administrator to demonstrate creating the first few patterns using the blocks themselves and for the participant to mimic that pattern. The physical component of this subtest would make it tricky to administer online, however, some previous researchers have tested it in comparison to in-person methods. Previous research shows that participants without cognitive impairments obtained similar scores between both methods of testing on the Block Design subtest (Mahon et al., 2021). Mahon et al., 2021 sent out the blocks required for this test along with other physical materials needed for the assessment in tamper proof courier bags prior to the assessment date and participants returned them via the same method. Moreover, in a study using stroke patients, Chapman et al., 2019, found no significant difference in participants' scores on the Block Design subtest between in-person and teleneuropsychological assessment methods, with both means being M = 36.14. Participants were provided with a research assistant while completing their teleneuropsychological assessment, who brought the blocks, materials and handled them during the assessment (Chapman et al., 2019).

Moreover, five more subtests from the WAIS-IV, including, Symbol Search, Visual Puzzles, Information and Coding have also shown to elicit similar scores from participants without cognitive impairments, across in-person and teleneuropsychological assessment methods (Mahon et al., 2021). Therefore, suggesting that participants' performance on these tests are not impacted by the assessment methods.

Word Knowledge and Verbal Concept Formation

Next, the Vocabulary subtest from the WAIS-IV is a measure of word knowledge, verbal concept formation and their ability to easily express ideas (Groth-Marnat & Wright, 2016). It is also a measure of people's long-term memory due to their need to search their memory to generate responses and explanations Groth-Marnat & Wright, 2016). Previous research has found no significant differences in participants' performance on the Vocabulary subtest between in-person and teleneuropsychological methods of assessment for both participants with and without cognitive impairments and demyelinating disorders (Svenn et al., 2003; Marra et al., 2020; Mahon et al., 2021). One study who did a comparison in a participant pool of children also found that their scores did not differ on this subtest across methods of testing (Harder et al., 2019). In the studies by Harder et al. (2019) and Mahon et al. (2021) the teleneuropsychological assessment was conducted in the participants' homes.

Overall, previous research suggests that people's performance on a lot of the WAIS-IV subtests are not impacted by the change to teleneuropsychological administration methods. However, thinking about these tests and their typical administration methods in-person, they are tests that could be easily transferred into teleneuropsychological methods of testing. Moreover, their administration does not require a lot of materials or speaking and they mostly just require the participant to go through the subtest themselves.

Executive Functioning

Another cognitive function assessed by neuropsychologists is executive functioning. The Verbal Fluency subtest from the Delis-Kaplan Executive Function System (DKEF-S) and is a measure of executive functioning but also has a component of short-term memory involved as participants must keep in mind what words they have already said (Groth-Marnat & Wright, 2016). Previous literature has found no difference between participants performance on this subtest between in-person and teleneuropsychological methods of assessment (Cullum et al., 2014; Harder et al., 2020; Gnassounou et al., 2022; Aguilar & Leguizamon, 2021). These findings are supported by another study by Parks et al. (2021), who also concluded that in-home teleneuropsychology assessment using the Verbal Fluency subtest, was a valid method of testing compared to face-to-face assessment. They did find significant differences in all scores between participants with and without neurocognitive disorders in the teleneuropsychological assessment condition but not for the in-person assessment condition (Parks et al., 2021). This score difference between participant groups does suggest that perhaps teleneuropsychology assessments may have more barriers depending on the condition of the client. The Verbal Fluency subtest contains two tests within, one being Letter Fluency. Looking specifically at Letter Fluency, previous research suggests there are no significant differences in participants' performance on this subtest in adults who had cognitive complaints, no cognitive impairments, cognitive impairments, dementia and stroke survivors, when comparing between in-person and teleneuropsychological methods of testing (Gnassounou et al., 2022; Alegret et al., 2021;

Chapman et al., 2019; Wadsworth et al., 2016; Cullum et al., 2014). Moreover, the same results were found in a pediatric population (children) (Harder et al., 2019).

The second test within Verbal Fluency is Category Fluency, similar to Letter Fluency but requires participants to recall words from a specific category. Previous literature suggests that there are no significant differences in participants scores on this subtest between in-person and teleneuropsychological assessment methods in adults with cognitive complaints, no cognitive impairments, cognitive impairments, dementia and stroke survivors (Gnassounou et al., 2022; Alegret et al., 2021; Chapman et al., 2019; Wadsworth et al., 2016; Brearly et al., 2017; Cullum et al., 2014). The same result was also seen in a pediatric population (Harder et al., 2020). However, one study did report finding a significant difference in participants' scores on Category Fluency between in-person and teleneuropsychological methods of testing but concluded that the effect size was small and the result was not clinically meaningful (Wadsworth et al., 2018). Marra et al., 2020 also noted that Category Fluency only showed moderate support for teleneuropsychological assessment validity, suggesting perhaps there could be variability in people's performance between the methods.

Next, the trial making test from the Delis-Kaplan Executive Function System (DKEF-S) that measures processing speed and executive functioning. It is a test that is commonly used to assess patients who have had a stroke, brain injury or cognitive impairments such as Alzheimer's or dementia. Previous literature found no significant differences on participants' performance on this subtest between in-person and teleneuropsychological methods of testing, in both participants with cognitive complaints and a range of cognitive impairments and clinically significant disorders (Gnassounou et al., 2022; Parks et al., 2021).

Short Term/Working Memory

Now focusing on the impacts of teleneuropsychological assessment on short-term memory, the Digit Span test is a subtest from the WAIS-IV that measures immediate verbal recall. Previous research shows that older adults with memory impairments, aged 60-80 years old had no significant differences in their performance on the Digit Span test (forwards and backwards) between face-to-face and teleneuropsychological assessment conditions (Gnassounou et al., 2022). Digit Span is both a measure of short-term memory and working memory. However, the teleneuropsychology assessments in this study were completed at the memory clinic or hospital on a MacBook computer rather than at their own home. Another study by Mahon et al., 2021 found a strong positive correlation (0.96) between in-person and teleneuropsychological methods of testing for Digit Span in healthy university students aged 18-40 years. In this study the participants completed the teleneuropsychology assessments in their own home and were delivered the materials required for the study. A similar study conducted by Aguilar & Leguizamon (2021) also found no significant differences between the methods of testing for both Digit Span forward and backwards. Overall, it seems that Digit Span is reliable in both methods of testing, however, in a survey at the end of the study by Aguilar & Leguizamon (2021) participants indicated that they preferred face-to-face methods so they could interact with the assessor without delays or audio problems. Moreover, participants expressed that they had worries around teleneuropsychological assessments that impacted their performance, such as, connectivity, no physical eye contact, no touching the materials and the added stress of seeing themselves on screen while completing tests (Aguilar & Leguizamon, 2021). This suggests that there may still be barriers for participants when using teleneuropsychological assessment methods.

Next, previous research has also looked at the Hopkins Verbal Learning Test Revised (HVLT-R) that measures immediate memory span and found no significant differences in participants' performance on this subtest between teleneuropsychological assessment and in-person assessment (Cullum et al., 2014; Parks et al., 2021). This result has been found for adults with and without cognitive impairments and stroke survivors on this test when comparing between assessment conditions (Alegret et al., 2021; Chapman et al., 2019; Wadsworth et al., 2018). Similar findings are reported in the pediatric population (Harder et al., 2020). Moreover, a review by Marra et al., 2020 found that this subtest shows strong support of teleneuropsychological assessment methods.

Lastly, the picture naming test is another test used to assess memory functioning and patients with cognitive complaints. Previous research has found no significant difference in participants' performance on this test between in-person and teleneuropsychological methods of assessment (Gnassounou et al., 2022).

Long Term Memory

Moving into the impacts of the change in assessment conditions on people's long-term memory functioning, the majority of the short-term memory tasks have a delay trial which provides information on a person's long-term memory functioning. It would be expected that long-term memory may not be affected by the change to teleneuropsychological assessment as people are retrieving already previously stored information from their long-term memory for these tests. However, if the assessment disrupted a person's ability on the short-term memory tasks, then it would be expected it would impact their performance on the delayed tasks and ability to recall information later for those tests. Moreover, because the literature suggests peoples' performance on the short-term memory tasks are not significantly different between in-person and teleneuropsychological assessment methods, the assumption could be made based on the memory model that their performance on the delayed trials would not be affected either.

Stand-alone Cognitive Assessments

This next section covers a few stand-alone assessments that are typically used to assess people for cognitive impairments such as Alzheimer's and Dementia. Firstly, the Mini Mental State Examination (MMSE) is a short screening tool commonly used to assess Dementia, Alzheimer's and other cognitive impairments (Molloy & Standish, 1997). Previous research has found no significant differences on participants performance on this test when comparing between in-person and teleneuropsychological assessment methods in both participants with and without cognitive impairments (Gnassounou et al., 2022; Parks et al., 2021; Cullum et al., 2014 & 2006)

Secondly, the Boston Naming test is a stand-alone test that measures word retrieval people's naming ability. Moreover, their knowledge of item names (e.g. they will be shown a picture of a tree and have to select from four answers, what the correct name for that picture is). It is a common test to assess aphasia in stroke patients (Roth, 2011). Previous literature has found that participants who had no cognitive impairments, cognitive impairments or in a clinical population had no significant differences between scores on this test between in-person and teleneuropsychological methods of testing (Parks et al., 2021; Stead & Vinson et al., 2019; Cullum et al., 2006; Cullum et al., 2014; Marra et al., 2020; Wadsworth et al., 2018).

Lastly, the Clock Drawing test is another stand-alone test that is commonly used as a screening tool for cognitive impairments and dementia as it measures peoples' understanding and visuo-spatial abilities (Agrell & Dehlin, 1998). Previous literature suggests there are no significant differences in participants' performance on this subtest between in-person and teleneuropsychological methods for participants with and without cognitive impairments and participants in a clinical population (Parks et al., 2021; Marra et al., 2020; Wadsworth et al., 2018 & 2016; Cullum et al., 2014 & 2006).

Overall, the literature comparing people's performance on different tests between in-person and teleneuropsychological assessment methods suggests that people's performance is not influenced by the change to remote assessments. However, as shown in Table 1, which is a summary of the previous literature, the studies vary in their sample population, tests used and time waited between tests. Therefore, there may still be reason to believe that the move to remote assessments may influence participants' performance.

Table 1

Summary of Previous Literature on the Comparison of Teleneuropsychological and In-Person Assessment Methods

Citation	Sample	Measures Compared	Test Interval	Key Findings
Wadsworth et al., 2018	197 Older Participants with cognitive impairment (n = 78) & without cognitive impairment (n = 119)	Boston Naming Test, Category Fluency, Clock Drawing Test, Digit Span Forward, Digit Span Backward, Hopkins Verbal Learning Test-Revised (HVLT-R) & Letter Fluency.	Same Day – 2.5hr Break	Significant difference between cognitively impaired and cognitively normal groups. Non-significant effect of administration condition Slight but statistically significant difference in category fluency (small effect size).
Marra et al., 2020	19 articles with 930 healthy participants, 359 with memory disorders, 28 with movement disorders, 78 with stroke accident, 30 psychiatric patients, 34 mixed clinical groups. Average age 65 or greater for each article.	Boston Naming Test, Category Fluency, Clock Drawing Test, Digit Span Forwards, Digit Span Backwards, Digit Span Total, Hopkins Verbal Learning Test-Revised, Letter Fluency, Matrix Reasoning & Vocabulary from WAIS-III.	Range between same day – 6/8 months	Letter fluency showed excellent support for TNP validity, Category fluency only moderate, Verbal learning showed strong support, Digit span (F,B,S) showed moderate-strong support. No difference in scores on the Vocabulary test between methods.
Wadsworth et al., 2016	84 participants with a mean age of 64. Mild cognitive impairment or dementia (n = 29), cognitively normal (n = 55).	Clock Drawing, Digit Span Forwards, Digit Span Backwards, HVLT-R, Letter Fluency, Category Fluency.	Same day (at least 20 mins apart). 2 participants retested 7-14 days later.	Small but statistically significant difference on Digit Span between remote and face-to-face mean scores. No significant differences between conditions for the other tests.
Gnassounou et al., 2022	150 participants aged 60-80 years old from memory clinics with a cognitive complaint.	Category Fluency, Digit Span, Letter Fluency, Mini-Mental State Examination (MMSE), Oral Picture Naming Test, Trial Making Test	4 months	No significant differences between assessment methods for each of the tests.

Cullum et al., 2014	202 adult subjects (mean age = 68), 83 with cognitive impairment, 119 healthy controls.	Boston Naming Test, Category Fluency, Clock Drawing Test, Digit Span Forwards, Digit Span Backwards, HVLT-R, Letter Fluency, MMSE.	Same day	No significant differences between the assessment methods for each of the tests.
Vahia et al., 2015	Participants aged 65 and older (No severe medical illness, psychiatric disorder, sensory impairments or previous neurological impairments)	-	-	-
Parks et al., 2021	111 in home (remote, mean age = 58) and 120 in person patients (mean age = 61). Patients had vascular disease/stroke, Alzheimer's, Parkinson's, multiple sclerosis, epilepsy, brain injury, cancer, frontotemporal dementia, psychiatric disorder, pain and sleep disorders and others.	Boston Naming Test, Calibrated Ideational Fluency Assessment, Clock Drawing Test, Digit Span Forwards, Digit Span backwards, HVLT-R, MMSE, Oral Trial Making Test,	-	No significant differences between the assessment methods for each of the tests.
Harder et al., 2020	58 participants, aged **6-20** years old, recruited through speciality programs for pediatric demyelinating disorders	California Verbal Learning Test Children (CVLT-C), CVLT-II, Delis-Kaplan Executive Function System (DKEFS) – Verbal Fluency (letter and category), Vocabulary, Digit Span Forwards, Digit Span Backwards (WISC-V, WAIS-IV), Category Fluency, Letter Fluency.	1-50 days	No significant results between the assessment methods for each of the tests.
Chapman et al., 2019	48 participants, mean age of 64 were recruited through a stroke survivor database.	Digit Span Forwards, Digit Span Backwards, Digit Span Total, Letter Fluency, Category Fluency, Hopkins Verbal Learning Immediate, Hopkins Verbal Learning Delay, Similarities	Average – 15.8 days apart	Participants did not perform systematically better in one assessment method compared to the other. Participants did perform worse on the HVLT-R test in the videoconference condition.
Alegret et al., 2021	338 participants assessed home-to-home were compared with 7990 participants assessed face-to-face (cognitively healthy, mild cognitive impairment, mild dementia)	Verbal Learning, Letter Fluency, Category Fluency, Similarities, Digit Span Forwards, Digit Span Backwards.	-	MCI participants had significantly better scores in all the tests in the home-to-home procedure – apart from Letter Fluency and Digit Span Forwards.
Brearly et al., 2017	Adult participants (17 and older), in a total of 25 studies for review	Digit Span, Category Fluency	-	Differences between videoconference and in-person are not significant
Grosch et al., 2015	8 participants from ages 67 to 85 with clinical diagnoses of depression, anxiety, schizophrenia and PTSD	Digit Span	5 minutes	No significant differences between assessment conditions.

Mahon et al., 2022	30 healthy participants aged 18-40 years	Arithmetic, Block Design, Coding, Digit Span, Similarities, Information, Matrix Reasoning, visual , Visual Puzzles, Vocabulary.	Mean = 7 days	No significant differences between assessment conditions
Svenn et al., 2003	32 healthy participants, with a mean age of 36 years tested first by telecommunication and 34 years tested first in-person	Logical memory, vocabulary, digit span subtests	Same day	High correlation between the testing methods for all subtests scores. Significant differences found for logical memory
Stead & Vinson, 2019	27 participants, aged 18-89 recruited from a university and local community. No previous history of psychiatric illness or neurological impairment, or substance misuse.	HVLT-R, Digit Span, mini-mental state, Boston naming test	10-15 minute break	Reliability between the two assessment methods.
Cullum et al., 2006	14 participants with MCI, 19 participants with AD.	Boston Naming Test, Category Fluency, Clock Drawing Test, Digit Span, HVLT-R, MMSE, Letter Fluency, MMSE.	Back-to-back	Slightly higher means in the HVLT-R delayed reacal

Gaps & Issues

Taken together, the previous literature suggests that overall people's performance does not differ between in-person and teleneuropsychological assessment methods of testing. However, people did report worries around teleneuropsychological assessment, such as, connectivity, no physical eye contact, not physically engaging with the materials, and stress of seeing themselves on screen (Aguilar & Leguizamon, 2021). This suggests that there may be barriers to teleneuropsychological testing.

Teleneuropsychological testing may also prove to be harder because people are using more cognitive load, thinking about their performance, using different cognitive functioning, but also experiencing worries around online testing, thus leading to more fatigue. Cognitive load is the optimal amount of information people can hold in their working memory at one time, which is usually around 7 bits of information, give or take two (Lezak et al., 2004; Turner, 2020). Research suggests that using software like Zoom increases our cognitive load as a few factors, such as concentrating on the screen and limited non-verbal cues, take up our working memory leading us to feel drained afterwards (Spataro, 2020; Tuner, 2020). Other factors that capacitate our cognitive load while being on camera include, knowing that people can see you and essentially feeling like you are on stage being judged (Tuner, 2020; Aguilar & Leguizamon, 2021). During a teleneuropsychological assessment, these factors may add extra stress and anxiety along with seeing yourself on screen which increases consciousness of physical appearance and expressions, also taking up far more energy than an in-person assessment. Therefore, this increase in cognitive load may impact people's performance on teleneuropsychological assessments.

A major gap in the literature is that previous research has also only examined a limited number of tests, mostly intelligence subtests from the WAIS-IV which suggests that a wider range of tests need to be assessed to determine the effects of the change to teleneuropsychological methods of testing on people's other cognitive domains.

Most non-significant findings suggest that perhaps intelligence tests are not impacted by changes in testing methods to online because of the types of tasks people are asked to do for those tests. The majority of these tests are visual tests measuring people's non-verbal skills or require people to retrieve information and think abstractly from their long-term memory. In addition, the majority of the previous research comparing in-person and teleneuropsychological assessments also do not focus on the area of functioning being measured by the tests chosen. Moreover, they compare and assess whether people obtain different scores on these tests between the test environments but do not look specifically at what cognitive domains these tests are assessing and whether people's cognitive functioning is being impacted by the change in test environment as well.

Additionally, previous literature around teleneuropsychology amid the COVID-19 pandemic is sparse and methods of most studies have participants coming into clinics to complete the online assessments and not completing them from home (Gnassounou et al., 2022; Wadsworth et al., 2018 & 2016; Cullum et al., 2014 & 2006). Teleneuropsychology assessments from people's homes may yield different results for their performance as their home may be an area with more noise and distractions, unstable internet connection etc. However, studies by Mahon et al., 2022 and Harder et al., 2020 did test people from their own homes and found no significant differences in participants' performance between in-person and at-home teleneuropsychological assessments. However, research for home-based teleneuropsychological assessments is still limited which is a major gap in the literature as people during times in the future similar to COVID-19 may not have adequate transport to clinics for assessments. Previous literature has not examined if home-based teleneuropsychological assessment and in-clinic teleneuropsychological would produce different results in participants' performance. Thinking about controllable factors of the in-clinic teleneuropsychological assessments such as a quiet space, computers and equipment set up already, different environment, assistants on hand, reliable internet connection, no distractions, there is reason to believe there may be. Therefore, this thesis aims determine the effects on people's cognitive performance of home-based teleneuropsychological assessments in New Zealand, using tests that are commonly used in New Zealand. In the hopes that the results will help towards filling these gaps in the literature and produce results that inform and provide useful information to psychologists in New Zealand.

Systematic Review of the Literature on Face masks

However, amid the COVID-19 pandemic, more PPE equipment has been used when doing in-person assessments. Moreover, assessments are carried out with both the assessor and client wearing masks. Mask-wearing is also going against the standardised conditions of testing which may impact people's cognitive performance on the tests.

Previous literature suggests that there may be an effect of mask-wearing on neuropsychological test performance. A literature review by Marler & Ditton (2020) explored the impact of mask wearing on communication in health care. They found that mask-wearing impacted the ability to form good therapeutic relationships and rapport between the healthcare provider and client (Marler & Ditton, 2020). The absence of rapport led clients to feel more disconnected, lonely and healthcare providers less trustworthy which all contributes to clients experiencing more anxiety (Marler & Ditton, 2020). Masks cover half of people's faces which eliminates the potential for lip reading, making it harder to read non-verbal cues and facial expressions for comfort (Marler & Ditton, 2020). All these barriers may lead to miscommunication, not being able to hear or understand properly and therefore increase the levels of frustration experienced by the client. However, masks do not only pose barriers to clients, but clinicians also found it more difficult to show their empathy towards clients, reassure and affirm them which contributed to their difficulty building rapport with the client (Marler & Ditton, 2020). Additionally, on top of the noise that is sometimes experienced in healthcare settings, mask-wearing makes communication even more difficult which shows the importance of doing neuropsychological assessment in standardised conditions with limited noise and distractions.

Other studies have found negative effects of mask-wearing on communication and hearing in both people with and without hearing loss (Saunders et al., 2021). Participants were asked questions about their previous experiences communicating with someone wearing a mask compared to their communication with someone who was not wearing a mask. Results showed that overall participants found it either hard or much harder to understand and hear people in a range of different settings (Doctors' appointments, pharmacists, hospital appointments, shop assistants, people at work) and even reported it being more difficult to understand their own family and friends (Saunders et al., 2021). The conversations participants had with someone else who was wearing a mask, led them to feel less engaged and less connected with the talker, and the impacts were greatest in healthcare settings (Saunders et al., 2021). This suggests that these impacts would be felt in neuropsychological assessments with the assessor wearing a mask. Additionally, as Malter & Ditton (2020) reported, the use of face masks increased anxiety and stress and led to people feeling more frustrated, embarrassed and fatigued when both listening to someone wearing a facemask or being the speaker wearing the face mask (Saunders et al., 2021). Therefore, not only do face masks have negative effects for people with hearing loss but also for people without hearing loss.

Furthermore, healthcare providers and clients (with/without hearing loss) from a study by Lee et al. (2022) reported that they increased their listening effort when the person they were communicating with was wearing a mask compared to when they were not wearing a mask. Lee et al. (2022) found that mask wearing led to people having more difficulty understanding conversations and a greater listening effort. These difficulties caused by mask-wearing increased people's cognitive load which slowed down the efficiency in healthcare settings, decreased people’s interest in conversation and healthcare providers found it more difficult to connect and build rapport with clients (Lee et al., 2022).

We already know that mask-wearing can lead to barriers in conversation and the ability to connect with the person wearing a mask, leading to feelings of disconnection (Marler & Ditton, 2020; Saunders et al., 2021). However, this level of disconnection and uncertainty around people’s feelings and empathy because of the concealment of people's facial expressions has clinically relevant and substantial effects on people’s anxiety levels (Saint & Moscovitch, 2021). Mask wearing increases the uncertainty associated with communication and increases perceived negative judgment (Saint & Moscovitch, 2021). Ambiguous social cues are more likely to be interpreted negatively when someone is wearing a mask, as their facial expressions are hidden which makes it hard to interpret how they are feeling (Saint & Moscovitch, 2021). Someone with social anxiety may find going to neuropsychology assessments even more daunting when the assessor is wearing a mask and may perceive barriers throughout the session, thinking they are not performing well or the assessor is judging them. Marler & Ditton (2020) suggested that wearing a mask increased people’s anxiety levels, which means clients may not be performing at their optimal performance during assessments and it may show in their test performance.

This is also supported by Kastendieck et al. (2022) who performed an online study where participants viewed 24 videos of a person (expressor) either wearing a mask or not wearing a mask, with either happy or sad expressions. Participants rated the emotional expressions of the person in the video and indicated how close they felt to them. Results showed that when the person was wearing a mask, participants perceived a decrease in emotion intensity compared to when the person was not wearing a mask (Kastendieck et al., 2022). The level of closeness was also higher when the expressor was not wearing a mask and they also felt closer to the happy expressors over the sad expressors (Kastendieck et al., 2022). These results suggest that when the assessor is wearing a mask during neuropsychological assessments, the client may feel more disconnected to the assessor and also may perceive them to have less intensity in either happy or sad emotions. While supporting previous literature, these findings also highlight the ambiguity, uncertainty and anxiety within communication with masks.

In contrast with other studies, mask wearing by the assessor may not have an effect on people's performance on neuropsychological assessment. Lichenstein et al. (2022) compared participants' performance on face-to-face neuropsychological assessments pre-COVID-19 and during-COVID-19. Patients aged 5-19 years old referred for outpatient evaluation within a pediatric neuropsychology service completed a series of neuropsychological tests face-to-face with safety precautions (Lichenstein et al., 2022). The assessors wore PPE gear, gloves, face coverings/shields and masks. Both the assessor and client were in the same room, physically distanced and the clients used a computer screen to engage with the testing stimuli. There were no significant difference in scores on the Wechsler Intelligence Scale for Children (WISC-V), WAIS-IV (digit span forward), WISC-V, WAIS-IV (matrix reasoning) and the California Verbal Learning Test (CVLT) between the two assessment conditions (Lichenstein et al., 2022). However, the protective measures used are not typical for in-person assessments. The use of a computer screen for clients is not typical as usually administration methods would be the same as the standardised conditions for in-person with only the additions of physical distancing and PPE gear. The use of the computer screen eliminates interactions between the client and assessor and therefore, decreases barriers with communication.

Other literature indicates that participants' performance on neuropsychological tests would be impacted by the assessor wearing a mask during in-person assessments. Truong, Beck & Weber (2021) found that mask wearing has an impact on people's ability to recall spoken sentences. German participants aged between 20-37 years listened and watched a video of a German speaker saying 48 different sentences either with or without wearing a face mask. Participants watched the videos of the speaker in counterbalanced order for both sentences and mask/no mask conditions. The results indicated that participants recalled fewer words when the speaker was wearing a mask compared to no mask (Truong, Beck & Weber, 2021). This finding suggests that the use of face masks interferes with people's ability to recall words or sentences, which gives us reason to believe that people's performance on memory tests that require recall of verbal information may be more greatly impacted by the use of masks.

Overall, past literature suggests that there will be an effect of face masks on participants' performance on neuropsychological assessments. Moreover, leading to more negative interference and negatively impacting their overall performance. Research suggests that it is because masks tend to muffle sound and make it harder for people to understand what someone is saying, in other words, decreased speech intelligibility (Bottalico et al., 2020). Muffled sounds through a mask, coupled with the inability to read lips or examine facial expressions poses a big barrier to communication. The inability to hear and understand the examiner during a neuropsychological assessment leaves room for miscommunication, misunderstanding of instructions, additional anxiety and disconnection, to name a few. However, what we do not know is whether wearing face masks specifically impacts people's performance on neuropsychological tests in regard to test scores. Face masks have been researched in many communication scenarios and only once specifically in relation to neuropsychological testing. This study was done using children and therefore, the impacts are still unknown in adults (Lichenstein et al., 2022).

Memory

Taken together, literature suggests that teleneuropsychology assessments and wearing face masks both increase a person's cognitive load (Lee et al., 2022). The increase in cognitive load means that people have fewer cognitive resources to perform at optimal on the neuropsychological tests. Moreover, if a person's cognitive load is being increased, then we should expect that people's memory will be impacted, if they are using up their cognitive resources on other things such as anxiety around seeing themselves on screen, increased listening effort, connection and managing resources to name a few (Aguilar & Leguizamon, 2021).

Increased cognitive load leads to poorer performance based on the information processing model of memory. Our working memory has a limited storage capacity, so when that is overwhelmed and cognitive load increases then information gets lost and does not get transferred into long-term memory (Lezak et al., 2004). When we hear verbal information, we need to be able to encode that information for it to be able to be held in our minds and then transferred into short and long-term memory storage. Encoding is the process of either listening or looking at the information being presented to us with full attention. In addition, if people are becoming more fatigued during online assessments due to increased cognitive load, it would lead to people's poorer memory performance on tests.

Our memory is composed of different components such as short-term memory and working memory and long term memory. Short term memory is the temporary storage of small amounts of information over a few seconds (Baddley et al., 2015). Working memory is our ability to keep that small amount of information in mind, while simultaneously completing complex tasks (Baddley et al., 2015). Essentially, it may involve holding that information in our mind briefly and re-arranging that information to recall back (Baddley et al., 2015; Lezak et al., 2004). These first stages of memory require a lot of attention and cognitive resources to efficiently encode the information provided to us. Encoding is the process of initially learning the information (Lezak et al., 2004). Once information has passed these stages of memory transfers to the long-term memory component. Long term memory is the storage of information for long periods of time such as specific life events or facts (Baddley et al., 2015). Sometimes when people have an injury or impairment affecting their brain, different components of their memory can be affected. Neuropsychological assessments contain multiple different memory tests to determine how different components of a person's memory are functioning and how it may be affected by their condition. These assessments provide useful information on how that person is able to remember things in their everyday life and what parts will be greatly affected by their condition, providing useful information for interventions or support.

The different types of tests used to assess memory tap into different components of memory such as working memory, short-term memory and long-term memory. Working memory tests may include complex tasks known as dual tasks, where the client will be asked to engage in immediate processing tasks such as reading sentences while also maintaining information in mind for recall (Alloway et al., 2006). Short-term memory tasks may include the client immediately recalling a list of words that were presented (Alloway et al., 2006). Moreover, the material that needs to be remembered can be presented via different modalities, such as, verbally or visually which test a person's verbal memory, auditory memory and visuospatial memory to gather in-depth information about that person's memory functioning. We would expect that the change to assessment conditions would have the greatest impact on people's verbal working memory and verbal short-term memory performance as people are actively trying to rehearse something that has been read to them, while also working to decode that speech. We know from previous research that people find it difficult to accurately remember verbal information when the assessor is wearing a mask. This may be because working memory and short-term memory tasks require a lot of resources and attention to encode that information, so anything that is a distraction from that task, such as anxiety, ambiguity in social communication and increased listening effort which increases cognitive load, may impact test performance. If information does not get past the highly attention-dependent encoding stage then it will not be remembered and people will not be able to efficiently recall that information (Lezak et al., 2004). Memory is a major aspect of a person's functioning and impacts their ability to perform specific tasks, such as, remembering names, plans for the day, and taking medication to name a few (Alloway et al., 2006). That is why it is extremely important for these neuropsychological assessments to provide valid and reliable information to help support these individuals. Therefore, this thesis will specifically focus on neuropsychological tests measuring memory performance as it is suggested to be the aspect of cognitive functioning that may be greatly impacted by these assessment condition changes. Furthermore, as we expect people's verbal working memory and short term memory to be greatly affected by these assessment condition changes, the tests that we expect to see the greatest impact on people's performance are Verbal Learning and Digit Span subtests as they both involve working memory and short-term memory components.

Summary & Importance of Research

To date most studies in the literature have not shown that teleneuropsychological assessment methods, also known as remote testing impact people's performance on tests. A limitation to previous literature is that studies have largely been limited to evaluating the WAIS-IV and the subtests from that test battery. It is possible that other tests that require more cognitive resources, such as, verbal memory tests, may be negatively impacted by these assessment method changes.

Previous research has shown that wearing face masks make it difficult to develop rapport, recognise emotions and understand speech (Marler & Ditton, 2020; Bottalico et al., 2020; Kastendieck et al., 2022). The ambiguity in social communication caused from mask-wearing has also been shown to increase anxiety and listening effort which collectively puts strain on cognitive load (Lee et al., 2022; Saint & Moscovitch, 2021; Marler & Ditton, 2020). In addition, one study found that face masks led to poorer recall of a series of spoken sentences (Truong, Beck & Weber, 2021). As yet it is not clear if wearing face masks does negatively impact adults' performance on a range of neuropsychological assessments.

Memory tests require a lot of cognitive resources to be able to efficiently hold information in mind and manipulate it to feed back as a response. We also know from previous literature that teleneuropsychological assessments increase a person’s cognitive load the same as mask wearing (Aguilar & Leguizamon, 2021). Therefore, these changes in assessment conditions leading to increased pressure on cognitive load may affect a person’s working memory ability negatively which would then impact their short-term memory performance as they may be experiencing encoding issues. Anxiety also acts as another contributor to a person’s cognitive load and reduces people’s encoding and processing capacity of their verbal working memory (Darke, 2008). Therefore, it is suggested that the change to teleneuropsychological assessment methods and the use of face masks during assessment may negatively impact people’s working and short-term memory. It could be assumed that if information is not being encoded then people’s long-term memory would also be negatively impacted as the information would not reach the stage of being accurately processed and stored.

Although previous literature around teleneuropsychology suggest it to be a reliable method of testing and found no differences in people's performance compared to face-to-face testing for these subtests, there are still significant barriers that have been highlighted that could impact performance. Such as physical interaction with the materials, delays and audio issues, connectivity issues, confidence using online methods and computers, no physical eye contact, harder to form rapport and the stress from seeing self on screen. These barriers may be more significant if participants were completing the assessment in their own home, compared to coming into a clinic and doing it on the computer. Additionally, if participants were completing assessments in their own home, it means the assessor may not be able to monitor and record behaviour accurately, and limits the area the assessor can see while the participant is engaging in the assessment (Mahon et al., 2021). It also allows more room for interruptions or distractions if they did not have a quiet place to do the assessment. Although there are benefits to teleneuropsychological assessment, we[1] aim to further investigate the potential negative effects around it and determine its reliability in New Zealand. Furthermore, the current research on teleneuropsychology examines participants' scores on subtests with face value, so they are only comparing a score between in-person and teleneuropsychological assessment conditions and do not examine specifically what that score means in terms of the impact the change in test environment may have on people's cognitive functioning over different cognitive domains. Moreover, if these assessment method changes can alter people's performance due to limiting their cognitive functioning.

If we couple mask wearing with teleneuropsychological assessments we should see the greatest impact on people's performance as this condition has the greatest influence on a person's cognitive load. Taking the increased cognitive load of doing assessments online and

[1] The research in this thesis is my own. However, participant data was collected alongside another Master's and two Honours students. Therefore, I often use the word "we" in this thesis to reflect those facts. Elsewhere, I use word "we" to refer to what is known (or not known) in the wider scientific community.

wearing masks, coupled with frustrations and anxieties of poor communication we would expect poorer performance. Mask-wearing is also suggested to increase people's listening effort, which adds another component onto their cognitive load and working memory and suggests that verbal tests may be more affected and we may see people have worse performance on these tests (Giovanelli et al., 2021). We also know that face masks muffle people's speech and pose a barrier to communication in many healthcare settings and teleneuropsychological assessments cause anxiety around connectivity issues. Therefore, taken together we should see the greatest impact on people's performance in this condition.

Therefore, the overall aim of this study was to determine the effects of non-standardised assessment methods on people's memory performance on neuropsychological tests commonly used in New Zealand. This thesis is specifically interested in people's performance on memory tests as it has not been covered extensively in previous literature. More specifically this thesis will focus on verbal short-term memory and verbal working memory tasks as they require a lot of attention and cognitive resources and if people's cognitive load is being negatively impacted by the non-standardised assessment methods, we would expect to see negative impacts on people's performance on these tests. The short-term memory and working memory tests that are used in this thesis are also the most appropriate for a New Zealand cultural context and are the most commonly used for neuropsychological assessments in New Zealand.

This thesis aims to determine the impacts of face masks and teleneuropsychological testing on people's short-term and working memory performance on neuropsychological tests. We conducted an experiment to answer two separate questions: What is the impact of wearing a face mask on people's performance on neuropsychological memory tests? What is the impact of teleneuropsychological assessments on people's performance on neuropsychological memory tests? More specifically:

- If the use of surgical face masks increases a person's cognitive load by greater listening effort, then we would expect people to have worse performance on neuropsychological tests when the assessor is wearing a mask compared to no mask.
- If teleneuropsychological assessment increases a person's cognitive load by additional assessment worries, then we would expect people to have worse performance on tests administered online compared to in-person.
- If both the use of surgical face masks and teleneuropsychological assessments increase a person's cognitive load by greater listening effort and additional worries, then we would expect people who complete the assessment via teleneuropsychological assessment with the assessor wearing a mask to have the worst performance compared to all other conditions.

Overall, this research is important to provide neuropsychologists with accurate information regarding the efficacy and validity of teleneuropsychological assessments or assessments with the assessor wearing a mask. More specifically this research will provide meaningful information on the impacts of face masks and teleneuropsychological assessments on people's short-term and working memory.

Method

Ethics

This study received ethical approval from the University of Waikato's Human Research Ethics Committee (Reference number: HREC 2022#04).

Design

Our study employed a 2 x 2 mixed method design, with test environment (Zoom, in-person) being a between subjects factor and masks (researcher wearing mask, researcher not wearing mask) being a within subjects factor. Participants were assigned to either zoom or in-person assessment conditions depending on preference, and they then completed the tests twice in a counterbalanced order (examiner with/without mask when delivering test instructions). As participants were required to complete two assessments between 1-2 weeks apart, alternate forms of the neuropsychological tests were used where possible to avoid any practice effects (Form A, Test manual A; Form B, Test manual B).

Participants

We recruited participants from the University of Waikato's Introduction to Psychology Research Program, who received 5% course credit (10% total) for participating in each assessment (See appendix A for flyer). Participants were also recruited using advertisements in the University webpage database and through online community pages, who received a $20 supermarket voucher ($40 total) for participating in each assessment (See appendix B for flyer). We aimed to recruit 30 participants for each assessment condition, either online or in-person, resulting in a subject pool of 60 participants at minimum. Our final sample consisted of 64 participants, 32 participants in the zoom condition and 32 participants in the in-person condition. Participants demographic information was compared between the zoom and in-person conditions for age, gender, ethnicity and language and displayed in Table 2. Comparisons were made using chi-square and *t*-tests.

Table 2

Comparison of Participant Demographics Between the Zoom and In-person Conditions

	Zoom		**In Person**		**Chi-Square Test/ *t*-test**
	n	%	*n*	%	(X^2, sig) (*t*, sig, cohen's *d*)
Age					
18-25	18	29	20	32.3	
26-35	4	6.5	7	11.3	
36-45	3	4.8	3	4.8	$X^2 = (4, N = 62) = 2.20, p = .700$
46-55	2	3.2	1	1.6	
56-65	3	4.8	1	1.6	
Age (*M*, *SD*)	30.17 (12.61)		26.03 (9.70)		$t(60) = 1.45, p = .150, d = .37$
Gender					
Female	20	32.3	22	35.5	
Male	10	16.1	9	14.4	$X^2 = (2, N = 62) = 1.08, p = .581$
Self-Describe	0	0.0	1	1.6	
Ethnicity					
NZ European	23	37.1	16	25.8	
Māori	4	6.5	13	21.0	$X^2 = (2, N = 62) = 5.96, p = .051$
Other	3	4.8	3	4.8	
Language					
English	28	45.2	30	48.4	
Māori	1	1.6	0	0.0	$X^2 = (2, N = 62) = 1.34, p = .512$
Other	1	1.6	2	3.2	

Note. N =64 (n = 32 for each condition). The reason *N* = 62 in this table is because two participants only disclosed their age and gender, but no other information (1 female, aged 23 from the zoom group and 1 male, aged 25 years, from the in-person group). Chi-squared results show no significant differences between groups for each demographic variable. Ethnicity was almost at the significant level but $p > .05$.

The demographics are presented in Table 2, however, two participants did not disclose all their demographic information as they did not have to answer any questions if they did not feel comfortable to. Those two participants did indicate their age and gender which is detailed in the note from Table 2. Overall, Table 2 displays that there were no significant differences for participant age, gender, ethnicity and language between the zoom and in-person conditions ($p > .05$).

Participants were initially screened for psychiatric and neurological disorders before completing any assessments. The inclusion used is similar to Mahon et al. (2021) and as recommended in the test handbooks. The criteria included: 1) Not currently taking any medication that affects cognitive ability, 2) No current medical conditions that affect eyesight, hearing, sense of smell or touch, 3) No diagnosis of epilepsy or diabetes, 4) No diagnosis of psychiatric disorder, 5) No history of a stroke or other neurological condition, 6) No history of a traumatic brain injury that resulted in unconsciousness for more than 20 minutes, 7) No other health conditions that may impact test performance, 8) Primary language is English, 9) provided written consent (See appendix C for screening questionnaire).

Materials

After giving verbal and written consent at the beginning of the assessment, participants were asked to complete a demographic questionnaire (Age, gender, ethnicity, first language, education and occupational training qualifications were covered) (See appendix F for demographic questionnaire). Participants were then assessed using a series of neuropsychological tests: Logical Memory, Visual Reproduction (Wechsler Memory Scale -IV (WMS-IV); Logical memory parallel form) (Wechsler, 2009; Schnabel, 2012), Vocabulary, Matrix Reasoning (Weschler Abbreviated Scale of Intelligence II; WASI-II) (Wechsler, 2011), Verbal Learning (Rey Auditory Verbal Learning Test (RAVLT); Crawford Verbal Learning Test (CAVLT)) (Schmidt, 1996; Crawford et al., 1989), Verbal Fluency (Delis Kaplan Executive Function System; DKEFS) (Delis et al., 2001); Digit Span, Letter Number Sequencing (Wechsler Adult Intelligence Scale -IV; WAIS-IV) (Wechsler, 2008), Spot the Word test (Baddley et al., 1992).

These tests were chosen with consultation with a registered neuropsychologist, taking into consideration the tests that are commonly used in New Zealand, the feasibility of administering the tests remotely, the assessment of cognitive domains most likely to be impacted by remote testing/mask use, the availability of parallel forms of the tests (to minimise practice effects) and to ensure the test session was not too long. The parallel forms of the tests were required as participants had to complete two assessments. These tests were arranged into two test manuals (Test manual A; Test manual B) and the order was counterbalanced between participants. The tests selected enabled estimates of IQ to be taken (WASI-II, Spot the Word) and concentration to be assessed (Digit Span). However, we did not calculate the estimates of IQ. Working memory (Letter-Number Sequencing), immediate, delayed and recognition memory (WMS-IV; Schnabel, 2012), learning (RAVLT, List A from Crawford; Crawford et al., 1989), fluency and cognitive flexibility (DKEFS) were also able to be measured. At the end of each session, participants were asked to answer questions about how they found the test sessions (using rating scales from Mahon et al., 2021) (See appendix G for the satisfaction questionnaire).

The materials for some of the tests chosen had to be adapted for online equivalent versions of the tests. PowerPoint slides were created for the online versions of the tests that required visual stimuli to be shown in full screen (Visual reproduction, matrix reasoning, vocabulary, verbal fluency). The examiner shared their screen during the assessment for each of these tests and clicked through the PowerPoint slides, rather than from flicking through hard materials of the tests.

This thesis will specifically focus on three of the neuropsychological tests: Verbal Learning (RAVLT, List A from Crawford), Verbal Fluency (DKEFS) and Digit Span (WAIS-IV) (Schmidt, 1996; Crawford et al., 1989; Delis et al., 2001; Weschler, 2008). Looking below at Table 3, the specific neuropsychological tests used, what test battery they are from, what aspect of cognitive functioning they measure and what scores we used is outlined.

Table 3

Neuropsychological Tests Chosen from the Test Battery for Analysis.

Test Name	Test Battery	Measure	Scores Used
Verbal Learning	RAVLT, Crawford List A	Immediate Working Memory	• Trial 1 & 5 total recall • Trial 6 total recall • Delayed total recall
Digit Span	WAIS-IV	Immediate Working Memory	• Digit Span Forward (DSF) Scaled Score • Digit Span Backward (DSB) Scaled Score • Digit Span Sequencing (DSS) Scaled Score • Total Scaled Score
Verbal Fluency	D-KEFS	Executive Functioning	• Letter Fluency Scaled Score • Category Fluency Scaled Score

The Verbal Learning task assesses immediate verbal auditory memory, new learning, susceptibility to interference and recognition memory. It involved the examiner reading aloud a list of 15 words (List A) for five consecutive trials, and each trial was followed by a free-recall test. The examiner recorded what words were said by the participant which allowed for the identification of any words offered by the participant that was not included on the list (Intrusion). After the fifth trial, an interference list of 15 new words (List B) was presented, followed by a free-recall test of that list. Immediately afterwards, participants were asked to perform a delayed free-recall of List A. The words from both list A and list B were read by the examiner at a rate of one word per every 1.5 to 2 seconds for each trial. After a 20-30 minute delay, participants were asked to complete another free-recall of list A. Followed by a recognition trial containing words from both list A and list B and similar sounding words, where participants were asked to indicate whether the words were on list A or not. The scores from the immediate recall task included the total recall and number of intrusions for each of the five trials for List A, the interference list (List B) and the final trial 6 for list A. For the delayed free-recall task scores were obtained for total recall and the number of intrusions.

The recognition task scores obtained were total number correct, total number of false positives and recognition score (total number correct - total number of false positives). (Schmidt, 1996)

A parallel form of the RAVLT was used for the test manual B, to avoid practice effects. The Crawford Verbal Learning Test was used and was administered the same as the RAVLT verbal learning test but with different word lists (Crawford et al., 1989).

The Digit span test from the WAIS-IV includes three tasks: Digit Span Forward, Digit Span Backward and Digit Span Sequencing (Wechsler, 2008). All three tasks were administered to the participants to assess their immediate working memory.

The Digit Span Forward task involved the examiner reading aloud a sequence of numbers to the participant at a rate of one digit per second, dropping their voice slightly on the last digit in the sequence. The participant was then required to recall back the number sequence in the same order as read by the examiner. Each test item (sequence length) was composed of two trials that were read by the examiner. After a total score of zero on both trials of an item, the test was discontinued. The scores obtained by the digit span forward task was the total raw score, scaled score, and longest digit span forward score (longest sequence length recalled) (Wechsler, 2008).

The Digit Span Backwards task involved the examiner reading aloud a sequence of numbers to the participant at a rate of one digit per second, dropping their voice slightly on the last digit in the sequence. The participant was then required to recall back the number sequence in reverse order than what was read by the examiner. Each test item (sequence length) was composed of two trials that were read by the examiner. After a total score of zero on both trials of an item, the test was discontinued. The scores obtained by the Digit Span Backwards task was the total raw score, scaled score, and longest Digit Span Backwards score (longest sequence length recalled) (Wechsler, 2008).

The Digit Span Sequencing task involved the examiner reading aloud a sequence of numbers to the participant at a rate of one digit per second, dropping their voice slightly on the last digit in the sequence. The participant was then required to recall back the number sequence in ascending order. Each test item (sequence length) was composed of two trials that were read by the examiner. After a total score of zero on both trials of an item, the test was discontinued. The scores obtained for the Digit Span Sequencing task were the total raw score, scaled score, and the longest Digit Span Sequencing score (longest sequence length recalled) (Wechsler, 2008). After all three Digit Span tasks were completed, the overall raw score was calculated by adding the sum of the raw scores on each test, then converted to a scaled score. The same Digit Span tasks were used in both forms of test manuals (A and B) as there was no alternate form for the Digit Span task and due to the complexity of the task, no practice effects were expected to be found.

The Verbal Fluency task from the DKEFS assesses executive functioning but requires demand on people's cognitive load by having to remember what they previously said (Delis et al., 2001). The Verbal Fluency task involved participants completing three different trials, one for Letter Fluency, one for Category Fluency and one for Category Switching (Category Switching was administered but the scores were not analysed for this thesis).

For the Letter Fluency task the examiner said a letter and gave participants 60 seconds to say as many words as they could that began with that certain letter (e.g. A). There were three different letter trials administered for letter fluency. Three alternate letters were provided in the DKEFS test manual to use at the second session and avoid practice effects (e.g. Letter A in test manual A and Letter H in test manual B) (Delis et al., 2001). Words said were not accepted if they were names of people, or places, or numbers and they could not repeat a word twice that had the same meaning but with different endings (e.g. take, takes, taking). The scores taken from this subtest were the total set loss errors across the three letter trials (words that violated the criteria), total repetition errors across the three letter trials and total correct responses across the three letter trials. Scaled scores were also calculated for the total correct responses score on letter fluency.

For the Category Fluency task, the examiner said a category (e.g. animals) and gave the participants 60 seconds to say as many words as they could that belonged in that certain category. The Category Fluency test was administered twice, for two separate categories. Alternate categories were provided by the DKEFS test manual to be administered in the second session (e.g. animals in test manual A and clothing in test manual B) (Delis et al., 2001). Alternate forms of this test were used to eliminate the possibility that participants would remember the letters and categories from the first session and perform better on their second session. Essentially removing any practice effects that might be found. The scores taken from this subtest were the total set loss errors across the two category trials (words that violated the criteria), total repetition errors across the two category trials and total correct responses across the two category trials. The scaled score was also calculated for the total correct responses score on category fluency.

At the end of the assessment participants were asked a series of questions regarding their satisfaction with the assessment to determine based on participant ratings, if either zoom or in-person assessments were favoured more over the other 1) rate the audio quality, 2) rate the visual stimulus quality, 3) privacy, 4) comfort, 5) convenience, 6) experience using computers. They were also asked if they perceived any cultural barriers during the test session or any other difficulties or barriers during the test session (See appendix G).

Procedure

Participants who expressed interest in participating were contacted via email to arrange a time for a screening phone call to ensure their eligibility to take part (See appendix C). If participants were eligible to take part in the study they were asked whether they would prefer to do the assessments online or in-person and allocated an ID number to prevent their information and scores being linked to personally identify them. If participants chose to complete the assessments online via Zoom, they were sent an email that informed them that they would be sent a pack of study materials and instructed them to not open the study materials until they got on the Zoom session with the examiner. Participants were also emailed a Zoom link for the assessment. The pack of study materials were sent to them via post. The study material packs included an information sheet, consent form, response booklet for the Visual Reproduction subtest and a response test for spot the word subtest (1x of each material for the tests were included in two separate envelopes labelled A and B for sessions one and two), as well as a facemask). The online assessments were conducted on Zoom (online video conferencing platform) with participants and the examiner using either a laptop or computer with a front-facing camera. Participants were required to arrange a workspace in a private room where they would have no distractions. To match conditions to the in-person assessment both the examiner and participants were required to wear masks on Zoom, with the examiner removing the mask to read the instructions of the tests, in the no-mask condition (due to COVID-19 protocols participants had to wear a mask during both sessions in the in-person condition). The examiner kept the mask on the whole time during the mask condition. The examiner joined the zoom call 10 minutes prior to the start time of the assessment, to organise the materials for the online versions of the tests. Once the zoom assessment began, the examiner first clearly talked through the information sheet which explained the study and what participation would involve (See appendix D for the information sheet). If the participant gave both written and verbal consent then the examiner instructed the participant what envelope, A or B, to open and use for that session. During the assessment, the examiner was required to share their screen with the participant for the Visual Reproduction, Matrix

Reasoning, Vocabulary and Verbal Fluency subtests. All other subtests as a part of the battery used were conducted as normal.

The in-person assessments were conducted either at the University of Waikato in a quiet, well ventilated room with no distractions, or in a quiet place in participants' homes where it was convenient for them. Both the participant and the examiner were required to wear surgical face masks, with the examiner removing the mask to read the instructions, in the no-mask condition. At the beginning of the assessment the examiner first clearly talked through the information sheet about the study and what participation would involve, and only began the assessment if the participants verbal and written consent was given. The examiner kept the mask on throughout the entire assessment during the mask condition. The in-person assessments were carried the same as Zoom, however, all tests in their usual in-person procedures.

At the end of all assessments participants were asked questions for the satisfaction questionnaire and answers were noted down by the examiner. Participants were required to answer their satisfaction from a scale of 1-5 on the following aspects of the session; audio quality, visual stimulus quality, privacy, comfort and convenience. Participants were also asked whether they perceived any cultural barriers or difficulties during the session (See appendix G). Whether on zoom or in-person, participants were thanked for their time and course credit was given for university participants, and vouchers were given to the participants from the community. If participants completed the session in-person the vouchers were given to them at the end of the assessment, however, if the assessment was completed on zoom participants were sent the vouchers in the post.

Statistical Analysis

After both assessments were completed, the participant’s scores on each individual test were calculated and converted to standard or scaled scores as described in the relevant test manuals for that specific test. All participant's data were input into Qualtrics through surveys and exported into an SPSS file for analysis. The data scores collected from the tests were analysed using IBM SPSS Statistics for Mac, version 28.0.1.1.

Before completing any analyses, all outlier, normality and homogeneity checks were completed to ensure that our data met the analysis assumptions and they were the right analyses to be performed for our data.

For the main analysis, the group means for each of the four conditions were compared using a 2x2 mixed ANOVA (test environment: zoom/in-person; masks: mask/no mask) in SPSS for each of the three subtests and all of their trials (verbal learning, digit span and verbal fluency). Partial Eta Squared effect sizes were calculated for the main analysis results, with a small effect size being $\eta p^2 = 0.01$, medium effect size $\eta p^2 = 0.06$ and a large effect size $\eta p^2 = 0.14$. Post hoc tests in the form of two independent samples *t*-tests and two paired sample *t*-tests were used to test the significant results found. Post hoc tests were important to determine the specific differences between the two conditions and explore which condition participants performed significantly better or worse in. As four different *t*-tests were run, we applied a Bonferroni correction for multiple comparisons so any significant results from the post-hoc were considered to be statistically significant if they were .013 or less ($p > .013$). Cohen's *d* effect sizes were calculated for the *t*-test results with a small effect size being $d = 0.2$, medium effect size $d = 0.5$ and a large effect size $d = 0.8$.

Another 2x2 mixed ANOVA was conducted to compare participants' satisfaction ratings between each of the four conditions, exploring whether participants preferred one method of testing better than the others. Partial Eta Squared effect sizes were also calculated for these ANOVA results using the same size interpretation mentioned above.

Exploratory analyses were conducted next, using paired sample *t*-tests to compare the alternate forms of the verbal learning and verbal fluency subtests. This was to determine whether the alternate forms that were used were equivalent. Cohen's *d* effect sizes were used for these *t*-test results, using the same size interpretation mentioned above.

To ensure there was no evidence of practice effects, paired t-tests were run for each of the three tests to compare participant's scores between session 1 and session 2.

Results

The Effects of Test Environment and Mask-Wearing on Memory Performance

In order to determine the effects of remote assessments and mask-wearing on people's performance on memory tasks, we carried out a series of 2 (Test environment - Zoom/InPerson) x 2 (Mask-Wearing - Mask/NoMask) mixed ANOVAs. To determine first whether a 2x2 mixed ANOVA was appropriate for our data, we needed to check if our data met assumptions for this analysis. There were some outliers in our data, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. However, further investigation of these outliers suggested they would not impact our results if they were kept in the analyses. ANOVA is also a robust method of testing and, therefore, outliers were not removed from the analyses. The data was normally distributed, as assessed by Shapiro-Wilk's test of normality. We had over 50 participants in our sample (64 participants) so we used the normality plots to determine the distribution. On examining these plots, we found that our data was approximately normally distributed. There was homogeneity of variances (all $p > .05$). and covariances ($p > .001$), as assessed by Levene's test of homogeneity of variance and Box's M test, respectively. Therefore, assumptions were met and the 2x2 ANOVAs were conducted.

What are the Impacts of Test Environment and Mask-Wearing on Memory Performance?

Participants' scaled scores for Digit Span and Verbal Fluency and total scores for Verbal Learning were compared across assessment conditions, shown in Table 4. Overall, there were no significant interaction effects between test environment and mask-wearing for all except one of the subtests. There was a statistically significant interaction effect for the Verbal Learning subtest, Trial 5, $F(1,60) = 8.76$, $p = .004$, partial $\eta p^2 = .126$. This interaction effect had a large effect size, with people in the zoom/mask condition having the lowest performance score. The interaction is presented in Figure 1.

The effect of mask wearing differed across the four assessment conditions for verbal learning, trial 5 (VLT5). In the zoom conditions, participants in the no mask condition had higher scores than participants in the mask condition. Whereas, when the tests were administered in-person, the no mask condition resulted in lower scores on VLT5 than when the examiner wore a mask. The condition where participants had the highest scores on trial 5 was in the in-person/mask condition and the condition where participants had the lowest scores on trial 5 was in the zoom/mask condition.

To examine further the interaction effect seen on VLT5, we ran a series of independent samples *t*-tests and paired samples *t*-tests to determine what conditions participants performed significantly better or worse in across test environment and mask-wearing. Before running the *t*-tests, all assumptions were checked and were met. There was no significant difference between participants' scores on VLT5 when the assessor was wearing a mask, for zoom compared to in-person ($t(61) = -2.22$, $p = .030$, $d = -0.56$ 95% [-1.06, -0.05]). Nor was there a significant difference between participants' scores on VLT5 when the assessor was not wearing a mask, for zoom compared to in-person ($t(61) = -0.91$, $p = .368$, $d = -0.23$ 95% [-0.27, 0.72]). However, participants in the zoom condition obtained significantly higher scores on VLT5 when the assessor was not wearing a mask compared to when they were wearing a mask ($t(31) = -2.64$, $p = .013$, $d = -0.47$ 95% [-0.83, -0.10]). There were no significant differences on VLT5 between the mask and no mask condition for participants who completed the assessment in person ($t(30) = 1.42$, $p = .165$, $d = 0.26$ 95% [-0.11, 0.61]).

Moving on to main effects, as can be seen in Table 4, there were no significant main effect of mask-wearing on any of the cognitive subtest scores ($p > .05$). All effect sizes were small for each of the subtests ($\eta_p^2 < .06$).

Moreover, there were no significant main effect of test environment on any of the cognitive subtest scores ($p > .05$). However, Category Fluency showed a medium effect size with participants gaining higher scores when completing the assessment on Zoom compared to in person ($\eta_p^2 = .067$). All other effect sizes were small ($\eta_p^2 < .06$).

Table 4

Comparison of the Effects of Test Environment and Mask-Wearing on Participants' Test Scores

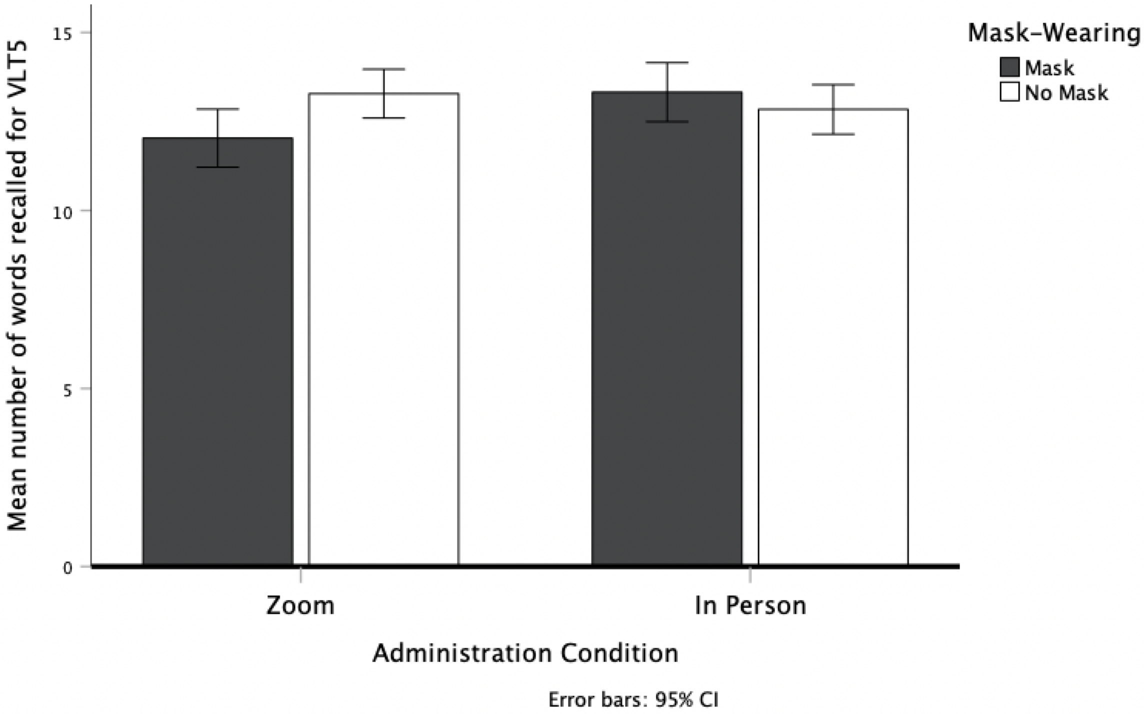
	Zoom (SD) *n* = 32		**In Person (SD) *n* = 32**		**Two way mixed ANOVA**		
	Mask	**No Mask**	**Mask**	**No Mask**	**Mask Wearing Effect (F, sig, partial eta square)**	**Assessment Condition Effect (F, sig, partial eta square)**	**Interaction Mask Wearing x Assessment Condition (F, sig, partial eta square)**
Verbal Fluency							
Letter Fluency	10.22 (3.01)	10.06 (2.69)	11.09 (3.29)	12.16 (5.41)	$F(1, 64) = 0.90, p = .348, \eta_p^2 = .014$	$F(1, 64) = 3.38, p = .071, \eta_p^2 = .052$	$F(1, 64) = 1.69, p = .203, \eta_p^2 = .025$
Category Fluency	12.25 (3.57)	11.50 (3.72)	10.56 (3.68)	9.94 (3.56)	$F(1, 64) = 2.02, p = .160, \eta_p^2 = .032$	$F(1, 64) = 4.46, p = .039, \eta_p^2 = .067$	$F(1, 64) = 0.02, p = .898, \eta_p^2 = .001$
Digit Span							
Digit Span Forwards	10.88 (3.10)	11.19 (3.43)	11.91 (3.86)	11.34 (3.69)	$F(1, 64) = 0.09, p = .771, \eta_p^2 = .001$	$F(1, 64) = 0.59, p = .445, \eta_p^2 = .009$	$F(1, 64) = 1.05, p = .309, \eta_p^2 = .017$
Digit Span Backwards	10.91 (2.92)	11.34 (3.44)	11.34 (3.03)	11.03 (3.03)	$F(1, 64) = 0.03, p = .865, \eta_p^2 = .001$	$F(1, 64) = 0.01, p = .928, \eta_p^2 = .001$	$F(1, 64) = 1.04, p = .311, \eta_p^2 = .017$
Digit Span Sequencing	10.56 (2.50)	10.69 (2.68)	10.41 (3.00)	10.16 (3.31)	$F(1, 64) = 0.03, p = .862, \eta_p^2 = .001$	$F(1, 64) = 0.30, p = .584, \eta_p^2 = .005$	$F(1, 64) = 0.27, p = .603, \eta_p^2 = .004$
Digit Span total	10.69 (2.80)	11.41 (3.06)	11.22 (3.11)	11.22 (3.43)	$F(1, 64) = 0.93, p = .339, \eta_p^2 = .015$	$F(1, 64) = 0.64, p = .801, \eta_p^2 = .001$	$F(1, 64) = 0.93, p = .339, \eta_p^2 = .015$
Verbal Learning							
Verbal Learning Trial 1	6.25 (2.64)	6.50 (1.63)	7.00 (2.22)	6.52 (2.13)	$F(1, 63) = 0.13, p = .718, \eta_p^2 = .002$	$F(1, 63) = 0.74, p = .394, \eta_p^2 = .012$	$F(1, 63) = 1.30, p = .259, \eta_p^2 = .021$
Verbal Learning Trial 2	9.34 (2.57)	9.34 (2.28)	10.00 (3.06)	9.29 (2.56)	$F(1, 63) = 1.09, p = .300, \eta_p^2 = .018$	$F(1, 63) = 0.28, p = .598, \eta_p^2 = .005$	$F(1, 63) = 1.09, p = .300, \eta_p^2 = .018$
Verbal Learning Trial 3	11.16 (2.64)	11.22 (1.90)	11.81 (2.74)	10.87 (2.91)	$F(1, 63) = 2.11, p = .152, \eta_p^2 = .033$	$F(1, 63) =. 0.07, p = .793, \eta_p^2 = .001$	$F(1, 63) = 2.75, p = .102, \eta_p^2 = .043$
Verbal Learning Trial 4	11.81 (2.51)	12.16 (2.90)	12.55 (2.13)	12.13 (2.08)	$F(1, 63) = 0.01, p = .916, \eta_p^2 = .001$	$F(1, 63) = 0.51, p = .480, \eta_p^2 = .008$	$F(1, 63) = 1.15, p = .288, \eta_p^2 = .018$
Verbal Learning Trial 5	12.03 (2.72)	13.28 (1.71)	13.32 (1.78)	12.84 (2.15)	$F(1, 63) = 1.71, p = .196, \eta_p^2 = .027$	$F(1, 63) = 0.89, p = .348, \eta_p^2 = .014$	$\mathbf{F(1, 63) = 8.76, p = .004, \eta_p^2 = .126}$
Verbal Learning List B	5.81 (1.75)	5.59 (1.98)	6.58 (2.06)	6.26 (2.00)	$F(1, 63) = 1.03, p = .313, \eta_p^2 = .017$	$F(1, 63) = 3.00, p = .088, \eta_p^2 = .047$	$F(1, 63) = 0.04, p = .846, \eta_p^2 = .001$
Verbal Learning Trial 6	10.78 (2.80)	11.56 (2.66)	11.68 (2.81)	11.68 (2.90)	$F(1, 63) = 1.25, p = .268, \eta_p^2 = .020$	$F(1, 63) = 0.68, p = .412, \eta_p^2 = .011$	$F(1, 63) = 1.25, p = .268, \eta_p^2 = .020$
Verbal Learning Delay	10.47 (3.04)	11.09 (3.48)	11.42 (2.98)	11.10 (2.96)	$F(1,63) = 0.15, p = .704, \eta_p^2 = .002$	$F(1, 63) = 0.49, p = .485, \eta_p^2 = .008$	$F(1, 63) = 1.43, p = .237, \eta_p^2 = .023$

Note. *n* = sample size.

Bolded results are significant. Data are presented as mean (SD) for each independent variable, with the main effects and interaction effects presented from the ANOVA.

Figure 1

Bar Graph Displaying the Significant Interaction Effect Between Test Environment and Mask-Wearing on Participants' Total Scores on VLT5.



Note. VLT5 = Verbal Learning Trial 5

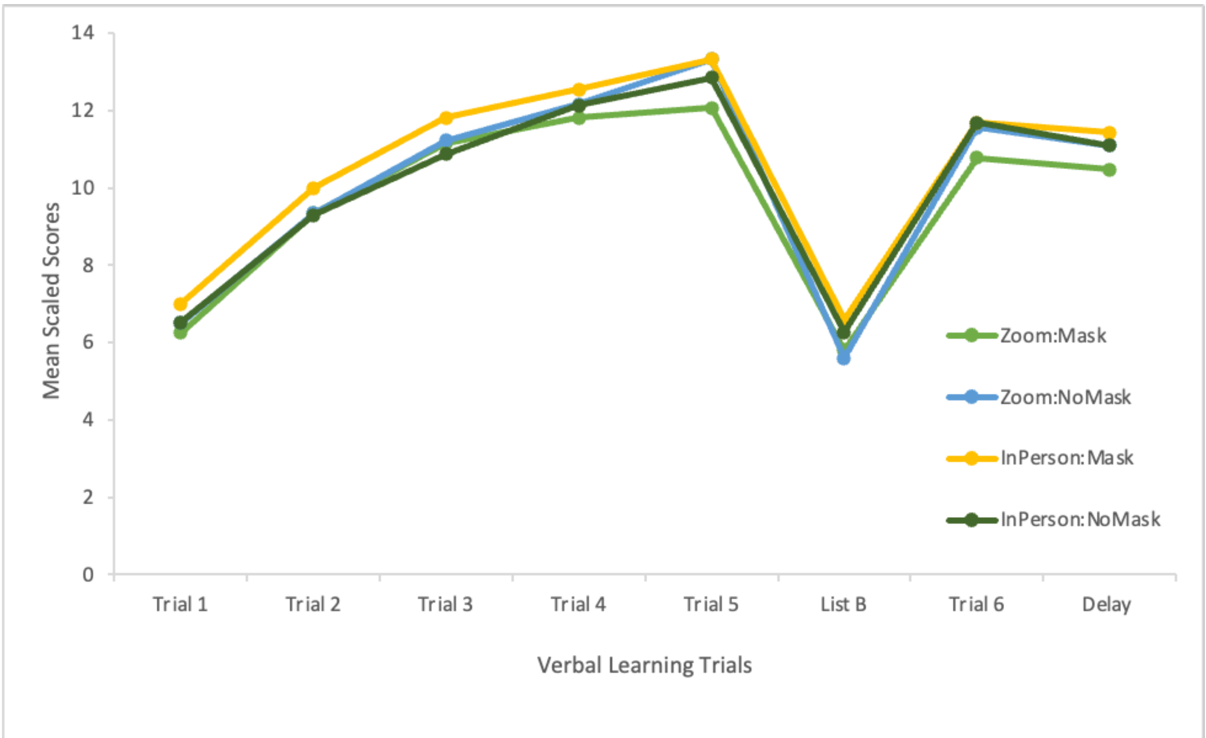
Verbal Learning (Learning effects)

Based on the significant interaction effect found for VLT5 when running the main analyses, participants' learning was examined to see whether there was a difference in learning across the four conditions. Learning across the trials in each condition was explored to examine whether participants were all experiencing increased learning or whether participants' learning was affected in the zoom/mask condition.

Figure 2 displays the scores obtained by participants over the 6 trials and delay trial on the Verbal Learning subtest, across each of the four conditions. Looking at Figure 2, it shows that there is effective increased learning, meaning, as participants go on through the trials they increasingly improve the number of words they recall. It is expected participants' scores would drop for the interference list B and for the delay. There also appeared to be no systematic differences in learning across conditions, which suggests that although participants in the zoom/mask condition recalled the least amount of words in trial 5, it was not due to the difference in learning across the conditions.

Figure 2

Participants' Learning for the Verbal Learning Subtest Across Each of the Conditions



The Effects of Assessment Condition and Mask Wearing on Satisfaction Ratings

To analyse participants' satisfaction ratings between each of the conditions, we carried out a 2x2 mixed ANOVA, to find out whether there were significant differences between participants' satisfaction ratings of the test session between each of the four conditions.

These data are summarised in Table 5. Overall, there were no significant interaction effects between participants' satisfaction ratings on any of the qualities between assessment condition and mask wearing ($p > .05$). There were no significant differences between participants' satisfaction ratings for audio, visual stimulus, privacy, comfort, convenience and confidence using computers for mask wearing compared to no mask ($p > .05$). There were also no significant differences between participants' satisfaction ratings for the qualities of the zoom assessments compared to in person assessments (*p* > .05) for all except audio quality and confidence with computers. Participants rated the audio quality to be poorer when they completed the assessment on zoom compared to in-person (*p* = .015). Participants also had higher ratings of confidence using computers in the zoom condition compared to the in-person condition (*p* = .030). The partial eta squared effect sizes ranged from small to medium. This suggests the amount of variance in participants satisfaction ratings that can be explained by the assessment conditions (mask/no mask, zoom/in person).

Table 5

Comparison of Participants Satisfaction Ratings Between the Four Different Conditions

	Zoom *n* = 26		**In Person *n* = 18**		**Two way mixed ANOVA**		
	Mask	**No Mask**	**Mask**	**No Mask**	**Mask Wearing Effect (*F*, sig, partial eta squared)**	**Assessment Condition Effect (*F*, sig, partial eta squared)**	**Interaction Effect Mask Wearing x Assessment Condition (*F*, sig, partial eta squared)**
Audio Quality	4.37 (0.69)	4.44 (0.70)	4.67 (0.97)	5.00 (0.00)	$F(1, 64) = 2.58, p = .115, ηp^2 = .057$	$\mathbf{F(1, 64) = 6.47, p = .015, ηp^2 = .131}$	$F(1, 64) = 1.05, p = .312, ηp^2 = .024$
Visual Stimulus Quality	4.78 (0.51)	4.78 (0.58)	4.78 (0.94)	5.00 (0.00)	$F(1, 64) = 0.85, p = .363, ηp^2 = .019$	$F(1, 64) = 0.68, p = .414, ηp^2 = .016$	$F(1, 64) = 0.85, p = .363, ηp^2 = .019$
Privacy	4.96 (0.19)	4.93 (0.27)	4.78 (0.94)	5.00 (0.00)	$F(1, 64) = 0.99, p = .325, ηp^2 = .023$	$F(1, 64) = 0.29, p = .593, ηp^2 = .007$	$F(1, 64) = 1.94, p = .171, ηp^2 = .043$
Comfort	4.81 (0.40)	4.93 (0.27)	4.67 (0.97)	4.89 (0.32)	$F(1, 64) = 2.61, p = .114, ηp^2 = .057$	$F(1, 64) = 0.58, p = .452, ηp^2 = .013$	$F(1, 64) = 0.29, p = .593, ηp^2 = .007$
Convenience	4.78 (0.58)	4.85 (0.36)	4.61 (0.98)	4.83 (0.38)	$F(1, 64) = 1.95, p = .170, ηp^2 = .043$	$F(1, 64) = 0.39, p = .535, ηp^2 = .009$	$F(1, 64) = 0.49, p = .489, ηp^2 = .011$
Confidence using computers	4.59 (0.69)	4.74 (0.66)	4.17 (0.92)	4.17 (0.86)	$F(1, 64) = 1.08, p = .304, ηp^2 = .025$	$\mathbf{F(1, 64) = 5.05, p = .030, ηp^2 = .105}$	$F(1, 64) = 1.08, p = .304, ηp^2 = .025$

Note. Participants voted based on a 5-point likert scale with 1 = very dissatisfied and 5 = very satisfactory
Data are presented as mean (SD) for each independent variable, with the main effects and interaction effect presented from the ANOVA.

Exploratory analyses

To ensure that our methods were robust and that our results were as accurate as possible, a series of repeated measures *t*-tests were run in SPSS to compare the alternate forms of tests used (Form A; Test manual A and Form B; Test manual B) to ensure that they were equivalent (for tests with alternate forms). Paired *t*-tests were also run to compare session 1 and 2, to determine whether there were any differences in participants' scores, indicating practice effects. We did use a counterbalanced design, so regardless if the alternate forms were equivalent or not, our results would not have been impacted. However, it was still interesting to check.

Comparison of alternate subtests forms between test manual A vs. B

Out of the three tests looked at in this thesis, only two have alternate forms, Verbal Learning and Verbal Fluency. Digit Span did not require an alternate form as it was highly unlikely participants would remember the digits from 1-2 weeks prior that would impact their performance on the second session.

Before running the paired samples *t*-test, assumptions were checked to determine if the *t*-test would be appropriate for the data. Multiple outliers were detected but were not extreme and they were kept in the analysis. All differences between participants' scores on the subtests between test manual A and B were approximately normally distributed, as assessed by visual inspection of a Normal Q-Q plot. The paired sample t-test was run normally, as the outliers were not believed to influence the results of the paired sample *t*-test.

The tests with alternate forms were Verbal Fluency (Letter and category fluency) and Verbal Learning. Paired sample *t*-tests were run in SPSS to compare participants' scaled scores on each of the subtests with form A compared to their scaled scores on each of the subtests with form B (see Table 6).

The paired *t*-test results showed that there were no significant differences between the alternate forms for Verbal Fluency, on both Letter and Category fluency ($p > .05$).

Moving on to the Verbal Learning test, the paired *t*-test results showed that there were significant differences between the alternate forms for all of the Verbal Learning trials apart from the interference list B trial ($p < .05$). Therefore, this suggests that the alternate forms of tests for Verbal Learning were not equivalent (Refer to Table 6). Moreover, Verbal Learning Trial 3 has a large effect size, with performance on test manual A being better than performance on test manual B. Therefore, suggesting that these results are of practical significance. Referring to Table 6, in each case, performance on the test manual A, which used the Rey Auditory Verbal Learning Test was better than performance on test manual B, which used the Crawford Verbal Learning Trial test.

Table 6

Comparison of Participants' Scores on Each of the Subtests Between Test Manuals A and B

	Form A mean score (SD)	**Form B mean score (SD)**	**Paired Samples *t*-test**	**Cohen's *d***
Verbal Fluency				
Letter Fluency	10.58 (3.14)	11.19 (4.38)	*t*(63) = -1.27, p = .208	*d* = -0.16 95% [-0.41, 0.09]
Category Fluency	11.09 (3.63)	11.03 (3.80)	*t*(63) = 0.13, *p* = .898	*d* = 0.02 95% [-0.23, 0.26]
Verbal Learning				
Verbal Learning Trial 1	7.22 (2.11)	5.90 (2.05)	***t*(62) = 4.77, *p* < .001**	*d* = 0.60 95% [0.33, 0.87]
Verbal Learning Trial 2	10.25 (2.63)	8.73 (2.38)	***t*(62) = 5.39, *p* <.001**	*d* = 0.68 95% [0.40, 0.95]
Verbal Learning Trial 3	12.03 (2.36)	10.49 (2.55)	***t*(62) = 6.41, *p* < .001**	*d* = 0.81 95% [0.52, 1.09]
Verbal Learning Trial 4	12.70 (2.54)	11.62 (2.17)	***t*(62) = 3.28, *p* = .002**	*d* = 0.41 95% [0.16, 0.67]
Verbal Learning Trial 5	13.49 (1.88)	12.24 (2.26)	***t*(62) = 4.62, *p* < .001**	*d* = 0.58 95% [0.31, 0.85]
Verbal Learning List B	6.13 (1.76)	5.98 (2.16)	*t*(62) = 0.54, *p* = .593	*d* = 0.07 95% [-0.18, 0.32]
Verbal Learning Trial 6	12.17 (2.43)	10.67 (2.93)	***t*(62) = 5.07, *p* < .001**	*d* = 0.64 95% [0.37, 0.91]
Verbal Learning Delay	11.95 (2.67)	10.08 (3.24)	***t*(62) = 5.86, *p* < .001**	*d* = 0.74 95% [0.46, 1.02]

Session 1 vs 2 (Practice effects)

All data assumptions were checked before running the paired samples *t*-test. The data was approximately normally distributed, analysed visually on a Q-Q plot and the paired sample *t*-test was run normally. Outliers in the data were not believed to influence the results of the paired sample *t*-test, whether they were included or not.

To determine if there were practice effects shown in participants' scores, we compared participants' scores for session 1 compared to session 2 with paired sample *t*-tests. The results from the paired *t*-test showed that there were no significant differences in participants' scores for each of the subtests between session 1 and 2 ($p > .05$). This means that there was no evidence of practice effects across sessions in our study (see Table 7) and all effect sizes were small ($d < .5$).

Table 7

Comparison of Participants' Scores on Each of the Subtests Between Session 1 and 2

	Session 1 mean score (SD)	Session 2 mean score (SD)	Paired Samples *t*-test	Cohen's *d*
Verbal Fluency				
Letter Fluency	10.55 (4.44)	11.22 (3.04)	$t(63) = -1.41, p = .164$	$d = -0.18$ 95% [-0.42, 0.72]
Category Fluency	11.06 (3.34)	11.06 (4.06)	$t(63) = 0.00, p = 1.00$	$d = 0.01$ 95% [-0.25, 0.25]
Digit Span				
Digit Span Forwards	11.14 (3.40)	11.52 (3.63)	$t(63) = -0.88, p = .381$	$d = -0.11$ 95% [-0.36, -0.14]
Digit Span Backwards	11.11 (3.08)	11.20 (3.11)	$t(63) = -0.26, p = .799$	$d = -0.03$ 95% [-0.28, 0.21]
Digit Span Sequencing	10.11 (2.76)	10.80 (2.93)	$t(63) = -1.99, p = .051$	$d = -0.03$ 95% [-0.50, 0.00]
Digit Span Total	10.78 (2.79)	11.48 (3.32)	$t(63) = -1.93, p = .059$	$d = -0.24$ 95% [-0.49, 0.01]
Verbal Learning				
Verbal Learning Trial 1	6.87 (2.29)	6.25 (2.02)	$t(62) = 1.97, p = .052$	$d = 0.25$ 95% [-0.00, 0.50]
Verbal Learning Trial 2	9.65 (2.75)	9.33 (2.48)	$t(62) = 0.93, p = .354$	$d = 0.12$ 95% [-0.13, 0.36]
Verbal Learning Trial 3	11.54 (2.55)	10.98 (2.57)	$t(62) = 1.84, p = .070$	$d = 0.23$ 95% [-0.02, 0.48]
Verbal Learning Trial 4	12.02 (2.68)	12.30 (2.13)	$t(62) = -0.81, p = .423$	$d = -0.10$ 95% [-0.35, 0.15]
Verbal Learning Trial 5	12.86 (2.24)	12.87 (2.11)	$t(62) = -0.05, p = .960$	$d = -0.01$ 95% [-0.25, 0.24]
Verbal Learning List B	6.05 (1.95)	6.06 (2.00)	$t(62) = -0.06, p = .953$	$d = -0.01$ 95% [-0.25, 0.24]
Verbal Learning Trial 6	11.16 (2.77)	11.68 (2.80)	$t(62) = -1.51, p =.137$	$d = -0.19$ 95% [-0.44, 0.06]
Verbal Learning Delay	11.02 (3.23)	11.02 (3.00)	$t(62) = 0.00, p = 1.00$	$d = 0.00$ 95% [-0.25, 0.25]

Discussion

The primary purpose of this study was to determine the impacts of face masks and teleneuropsychological assessment on people's working and short-term memory performance on neuropsychological tests. With a sample of 64 people, we found that the use of face masks and teleneuropsychological assessment did not impact people's performance on the Verbal Fluency, Digit Span or Verbal Learning subtests. Our first hypothesis that the use of surgical face masks increases a person's cognitive load and, therefore, decreases their performance compared to no mask, was not supported. We found no difference in participants' scores when the assessor was wearing a mask, compared to no mask on the Verbal Learning, Digit Span and Verbal Fluency subtests. Our second hypothesis that teleneuropsychological assessments increase a person's cognitive load and, therefore, decrease their performance compared to in-person assessments, was also not supported. We found no difference in participants' scores when they completed the assessment on Zoom compared to in-person. However, we did find an interaction effect on trial 5 of Verbal Learning, where people in the zoom/mask condition recalled the lowest number of words on that trial, compared to the other three conditions. People in the zoom/mask condition recalled significantly fewer words than people in the zoom/no mask condition. This finding partially supports our last hypothesis that when participants completed the assessment on Zoom, with the assessor wearing a surgical face mask, this condition would put the greatest amount of pressure on people's cognitive load and, therefore, these participants would have the worst performance on the tests compared to participants in the other three conditions. We only saw this significant interaction effect for Verbal Learning trial 5, where participants in the zoom/mask condition had the worst performance and recall on this trial.

Contrary to our hypotheses that teleneuropsychological assessment and mask-wearing would increase a person's cognitive load and, therefore, decrease their performance, we found that people's performance was not impacted by the change in assessment conditions.

An explanation for why we found no significant effects of teleneuropsychological assessment on people's working and short-term memory performance, may be because COVID-19 in New Zealand has been around for almost three years and majority of our participants were students who would have been through university attending lectures online through Zoom during the nation-wide lockdowns, so they may have become well equipped to adapt to Zoom audio quality, fatigue and anxieties now. Additionally, mask mandates have also been in New Zealand for a long period of time, so an explanation may be that people have gotten used to wearing and communicating through face masks, eliminating significant negative effects of them on their performance. Previous research has found that when people wear N95 and surgical face masks, they speak at a lower rate compared to no mask, to produce clearer speech, which suggests that people may try to compensate for the muffled speech that masks cause to improve their speech intelligibility (Magee et al., 2020; Yi et al., 2021). With assessors perhaps unconsciously speaking clearer to match how they would sound without a mask, it would explain why we saw no significant difference in participants' performance between the mask and no mask conditions.

In support of the interaction hypothesis, the results showed that participants in the zoom/mask condition recalled the fewest words on Verbal Learning trial 5 compared to the participants in the other conditions. This result may suggest that perhaps this trial combined with the assessment conditions, increased people's cognitive load the most and, therefore, impacted people's ability to recall many words.

However, we needed to determine whether this low score was a result of not remembering the words, or if it was a result of misheard words. During scoring, a word was scored as an intrusion word when the participant recalled a word that was not on the original list. Sometimes the word was wrong and other times it was a result of the participant mishearing a word that the assessor said when reading the list. Frequently, an error or misheard word that had arisen in an earlier trial would continue to appear throughout following trials, often in the same position in the list, which we termed to be misheard (Lezak et al., 2004). For example, participants often misheard the word “nail” from the Crawford Verbal Learning test, and recalled it as “mail”. When looking at the data, it was discovered that people in the zoom/mask condition on VLT5 had nine misheard words overall, compared to one misheard word in the in person/mask condition, three misheard words overall in the in person/no mask condition and three misheard words in the zoom/no mask condition. Therefore, overall people in the zoom/mask condition did mishear more words than the other conditions, which suggests that their significantly less words recalled could be due to their increased cognitive load and fatigue on trial 5. After completing four trials of Verbal Learning on Zoom, heading into the fifth trial, participants may be reaching a place of fatigue. We know that prolonged time on zoom can lead people to feel fatigued due to cognitive load pressures such as seeing yourself on screen, anxieties around connectivity and audio problems, no eye contact or physically touching materials (Aguilar & Leguizamon, 2021). This combined with the pressures of mask wearing such as increased listening effort, increased anxiety and disconnect may all contribute to the cognitive overload of participants leading to fatigue. A reason we may have only seen this result on trial 5 is that participants may have been at their most fatigued due to the repetition of the Verbal Learning task. Verbal learning is a task that continuously uses participants' attention and when people are forced to continuously use their attention, it increases their cognitive load (Barrouillet et al., 2007). The Verbal Learning subtest also requires people to continuously retrieve (recall) information from their memory, multiple times over multiple trials which again is putting more pressure on their cognitive load. Previous research suggests that increasing the amount or number of times that someone has to retrieve information, even only slightly, impacts their ability to recall that information accurately (Barrouillet et al., 2007). The fifth recall trial, coupled with the added stresses and fatigue caused by video conferencing methods, and difficulties of communication through masks, may have added the most pressure to participants' cognitive load in this condition (Shoshan & Wehrt, 2021).

Moreover, the interaction result seen on trial 5, supports the hypothesis that people would recall fewer words and perform worse when they were in the condition with the assessor wearing a mask and completing the assessment on Zoom. This was based on the cognitive load theory as this condition may have put the most pressure on participants' cognitive load, which impacted their working memory and, therefore, short-term memory. However, even though we have seen a statistically significant interaction result, the difference in means between the conditions is relatively small and would be equivalent to a difference in one or fewer words recalled (see table 4). The small difference suggests that our results are not clinically significant or meaningful as they were less than 1.5 standard deviations different (Harvey, 2012; Sharma, 2021). Moreover, the effect size was also small, which supports the conclusion the result is not clinically meaningful, even though it was statistically significant ($\eta p^2 = 0.1$). This result also may have been a spurious finding as a result of the multiple different analyses that were conducted, which increased the rate of type 1 errors (family wise error). A type 1 error is when it is concluded that a result is statistically significant or true, when it is not. This is plausible because there was no difference in learning across the three conditions when plotted in a graph, suggesting that people were not performing on the test much differently overall compared to the other conditions (Figure 2).

In addition, to determine whether the design and methodology of our study was sound, we performed exploratory analyses. We found no significant practice effects across session 1 and session 2 which means people's performance in the second session was not impacted by their experience on the first test session. This result means that there was no confounding effect of session on the participants memory performance, even for the tests without alternate forms.

We used alternate forms of Verbal Learning and Verbal Fluency subtests for test manual A and B to eliminate practice effects. To determine whether these alternate forms were equivalent we compared participants' scores on each of the subtests between test manual A and test manual B. We did not use an alternate form for Digit Span as it was unlikely that participants would have remembered the digit sequences from the first session, 1-2 weeks later. Our results showed that there was a significant difference in participants' scores between the alternate forms A and B on the Verbal Learning subtest. Suggesting that those alternate forms were not equivalent. Looking further, we found that people performed better and had a better recall on the Verbal Learning subtest with the word list from test manual A, which was the Rey Auditory Verbal Learning Task (RAVLT) (Schmidt, 1996). This being compared to the alternate form of the Verbal Learning subtest, the Crawford Auditory Verbal Learning Task (Crawford et al., 1989) in test manual B. However, this was only seen for trials 1-5 and trial 6. We did not find a significant difference in participants' scores on the interference list B meaning that people had similar recall performance on this list, regardless of the different word lists. Therefore, list B from the alternate form was equivalent to list B from the original test. We found no significant difference between the participants' scores on the Verbal Fluency subtest from the Delis-Kaplan Executive Function System (D-KEFS) (Delis et al., 2001), between the alternate forms used for each session. These alternate forms (letters and categories) were given by the original test (D-KEFS) to be used for re-test, so no significant differences were expected.

An interpretation of the result that the alternate forms used for Verbal Learning were not equivalent, suggests that the word list from the Crawford Verbal Learning task is a harder list of words for people to remember compared to the original RAVLT. The Crawford alternate form was developed in the United Kingdom (UK) with a group of 60 participants, who were split into two groups by putting people in pairs based on their age, sex, and years of education (Crawford et al., 1989). These two groups had no significant difference on IQ scores. One person in each pair was administered the original RAVLT and the other person administered the new word list. These words were matched to words from the original AVLT and chosen based on their frequency of occurrence in the English language (Throndike & Lorge, 1944), word length and serial position of a word in the original list (Crawford et al., 1989). However, the original RAVLT was developed in a group of French-speaking people and later developed into the English language and then other languages (Bean, 2011). Taking a closer look at the word lists, you can see that perhaps the words from the RAVLT are closer semantically related than the words in the Crawford list, making it easier to remember. For example, words such as “Garden, Hat, Farmer" or “School, Parent” compared to words from the Crawford that are not related in the same list positions “Letter, Bed, Machine” or “Heart, Desert”. Literature also states that we have shallow and deep levels of processing during the information encoding stage, and if we are using deeper processing for example, when something is more meaningful then it leads to better memory retention (Craik & Lockhart, 1972). Which might be happening if the words from RAVLT are more semantically related, making them more meaningful and leading to deeper processing and therefore, better memory retention and perhaps retrieval. This may be affecting the equivalence of these alternate forms.

However, because our study followed a counterbalanced design, this means our results should not have been impacted by the difference in our alternate forms. This is because participants completed both test manuals A and B, in a counterbalanced order, so not all participants completed A first or B first. Therefore, this was not a concern for our main analysis results. The average difference between participants’ scores on the Verbal Learning subtest between the RAVLT and Crawford, was 1.68 between all trials 1-6 (All differences added up, divided by 6 as list B and delayed trials were not included). All effect sizes were medium to large (0.5 - 0.8) with only one small (0.4). However, this is not a clinically significant result because the difference between participants' mean scores were less than 1.5 standard deviations different (Harvey, 2012; Sharma, 2021). Even though this result is not clinically significant it is still a meaningful difference. This result is important for neuropsychologists in New Zealand as it suggests that the RAVLT and Crawford Verbal Learning subtests cannot be substituted or compared as people score lower on the Crawford compared to the RAVLT. Therefore, neuropsychologists would need to be cautious when using alternate forms of Verbal Learning to ensure that they are equivalent and our results suggest that the RAVLT and CAVLT are not.

Implications of Results

Overall, our main findings that teleneuropsychological assessment and mask-wearing do not impact people's working and short-term memory performance on neuropsychological tests builds on the existing literature. Our study looked specifically at the impacts of teleneuropsychological assessment on memory performance which is the first study to do this in the literature. A study by Mahon et al. (2021) assessed the feasibility of at-home based teleneuropsychological assessment compared to in-person, using intelligence tests and a sample of people aged 18-40 with no cognitive impairments. Our results are similar to the results by Mahon et al. (2021), that there is no significant difference in participants' scores between conditions and, therefore, teleneuropsychological assessment is a feasible method of testing. Moreover, our findings contribute to a clearer understanding of how the change in assessment conditions impacts people's cognitive domains through the use of different tests. Like Mahon et al. (2021) we used an at-home-based teleneuropsychological assessment method, which was also used by one other study, Alegret et al. (2021). The inconsistency in the existing literature is highlighted as our results do not fit with the findings of Alegret et al. (2021), who found that people performed better on Verbal Learning, Verbal Fluency (Category) and Digit Span Backwards at-home compared to in-person. However, they used a sample of both people with and without cognitive impairments and people were readily available to assist with any difficulties during the at-home assessments. Our results provide a new understanding that perhaps assistance is not needed for at-home based teleneuropsychological testing, however, our results are only true for a population aged 18-59 who have no cognitive impairments. Therefore, it is unclear whether we would see the same results for people with cognitive impairments or for older people.

While previous literature has only focused on the impacts of teleneuropsychological assessments on participants' scores and different cognitive domains and tests, our results for Verbal Fluency and Digit Span were similar to those of Gnassounou et al. (2022) who found no significant differences in participants scores on these subtests. While we found the same results, they used a sample of older cognitively impaired individuals, which we would expect the change to remote testing would be harder but it contributes to the varying literature and combined with our results, suggests a wider range of people that might not be impacted by teleneuropsychology assessment. They also used a 4 month gap between assessments while we used a 1-2 week gap which strengthens the findings that teleneuropsychological assessments are a valid method of testing and suggest that the gap between assessments does not impact participants' scores as long as administration methods are sound to eliminate practice effects.

The majority of findings in the literature support that people's scores on the Digit Span subtest are not impacted by the change to remote assessments (Bearly et al., 2012). Our results strengthen those findings, however, we would have expected that with a test like Digit Span, that tests people's working memory and occupies their cognitive load, people would have performed worse on Zoom with all the added cognitive load pressures (Barrouillet et al., 2007). While we did not find evidence for this theory it may have been because the task was not cognitively demanding enough, or as mentioned earlier, peoples' adaptability to speaking with face masks on and communicating through Zoom.

While most research looks solely at the comparison of teleneuropsychological assessments to in-person assessments, we also looked at the impacts of wearing face masks during assessment on people's memory performance. We found that mask-wearing by the assessor also has no impact on a person's working or short-term memory performance compared to the assessor not wearing a mask. Our study was the first to look at the effects of mask-wearing on memory performance, which builds on the existing effects of mask-wearing in the literature.

Our results do not fit with previous findings on the negative impacts of mask-wearing on communication, anxiety levels, and cognitive load (Saunders et al., 2021; Kastendieck et al., 2022; Saint & Moscovitch, 2021; Lee et al., 2022; Gutz et al., 2022). These studies suggest that people find it harder to understand and hear people in healthcare settings through masks and feel less engaged and connected (Saunders et al., 2021). Moreover, our results do not support the theory that people increase their listening effort when they are communicating with someone wearing a mask, which in turn increases their cognitive load and leads them to have more difficulty in understanding conversations (Lee et al., 2022). Theory suggested that coupled with increased listening effort, people also have higher levels of anxiety because of their uncertainty around people's feelings, expressions and judgements which is all putting pressure on their cognitive load (Katendieck et al., 2022; Saint & Moscovitch, 2021). Contrary to the expectation that people's cognitive load is being pressured and they should have worse performance when the assessor was wearing a mask, we did not find evidence to support this theory.

Our findings do support the findings by Lichenstein et al. (2022) who also looked at the impact of mask wearing on teleneuropsychological test performance and found no significant differences between participants scores on Digit Span, California Verbal Learning test, and Matrix Reasoning compared to when the assessor was not wearing a mask. However, they used a sample of young patients who were assessed by an assessor wearing

PPE but also had a computer to complete the tests on in the same room (Lichenstein et al., 2022). We would expect that the presence of the computer would eliminate interactions between the client and assessor and would decrease any barriers in communication. However, our findings build on these and suggest that the presence of a computer does not impact people's performance when the assessor is wearing a mask or not.

Even though our findings do not fit with the cognitive load theory, we used only surgical masks in our study which has also been said to have the smallest influence on speech intelligibility compared to cloth masks and transparent masks (Brown et al., 2021). Therefore, our results may have been the impact of the type of mask we used. However, overall, they do suggest that the impact of mask wearing on people's memory performance is not significant as they do not influence cognitive load or muffle speech and communication enough to see an effect.

Our research provides a new insight into the relationship between teleneuropsychological assessment and mask-wearing on people's working and short-term memory performance. Our significant interaction result for trail 5 of Verbal Learning which shows that participants performed worse in the zoom/mask condition fits with findings by Chapman et al. (2019) who found that stroke survivors performed worse on Verbal Learning in the teleneuropsychological condition and findings of Truong, Beck & Weber (2021) who found that mask wearing impacts people's ability to recall spoken sentences. While our findings are seen only when the conditions are coupled together for this specific trial, our findings still provide insight into the relationship between these assessment conditions. Our study was the first to look at both of these conditions together and their specific impacts on memory. Although, as mentioned before, this result may be a spurious finding and is not clinically significant.

Even though results were not statistically significant and these assessment conditions do not alter participants working memory and short-term memory performance, there are other implications of mask-wearing and teleneuropsychological assessment that were discovered during participants' satisfaction rating questionnaires. We found that there may be significant cultural implications because of the disconnect between the assessor and client caused by the conditions. One participant from the zoom/mask condition reported that when both people were wearing masks, it acted as a cultural barrier. "Kanoki ei Kitea" which means face/seen and it is of cultural importance to Māori to see each other's faces. This particular participant also had a moko kauae – lip and chin tattoo received by Māori women (New Zealand Tourism, n.d) – that was hidden behind the mask, hiding her culture, community, whakapapa, status, abilities, where she was from, her history and making it harder to whakawhanaungatanga. Whakawhanaungatanga is an important process of relationship building for Māori, and is important in an assessment space to ease anxieties and build rapport. Moreover, she also mentioned that completing the assessment on Zoom alone was difficult. Although this was a comment from one participant, it has great significance in the New Zealand context as the feelings of cultural invisibility, anxieties around being in a westernised space, lack of rapport and trust and connection, and overlooking the role of cultural identity on assessment, can severely impact the results of the assessment (Dudley et al., 2014). Neuropsychologists need to be aware that when Māori come into the clinic space in New Zealand, they are entering a space that is foreign to them and whakawhanaungatanga, building trust and connection, as well as engaging in cultural practices of Māori is important to help ease anxieties and worries and for a valid and reliable test outcome. Not only does it allow for a valid and reliable outcome but it also ensures that they have a positive experience and feel adequate in the clinic space. Additionally, this participant mentioned that she did not know the meanings of some words during vocabulary and verbal learning. She mentioned that she was unsure if it was due to cultural differences of her being Māori and these tests and assessments being Westernised or whether it was relevant to world experience.

Moreover, even though we found no significant impacts of teleneuropsychological assessment or mask-wearing on participants' performance on Verbal Learning, Digit Span and Verbal Fluency, people did mention some negative experiences in these conditions. When participants completed the assessment on Zoom, they reported internet connectivity issues, audio cutting out, computer lags, wifi issues, sound volume too quiet, issues with lagging and causing words to be missed during Verbal Learning, which also meant that the audio and video were not in sync. Although we found no significant differences in participants' satisfaction ratings, the comments participants made in relation to teleneuropsychological assessment fit with the negative effects and concerns of these assessments commented by other participants in the literature (Aguilar & Leguizamon, 2021; Rochette et al., 2021).

Mask wearing participants commented that it was hard to lip read when the assessor was wearing a mask and when the assessment was completed on zoom as well, they mentioned that it was hard to read the assessor's body language. Participants also mentioned that on Verbal Learning, they experienced some glitches, mumbled words which was worse when the assessor was wearing a mask too. This supports the muffled speech sound that masks create (Bottalico et al., 2020).

Strengths and Limitations

One limitation of our study is that the majority of our participants were a student-aged population with only a few older adults, therefore, a critic may wonder if our results would still be seen in older populations or people who are cognitively impaired. It may be assumed that older adults and cognitively impaired individuals may be more impacted by the changes in assessment conditions. However, previous studies suggest that they would not be affected by the change to online assessments (Marra et al., 2020). Although, research on mask-wearing and the impacts on neuropsychological assessment performance is limited. Additionally, neuropsychological assessments in New Zealand are typically not used for minor issues so our young to middle-aged population are not typically the people who would be needing a neuropsychological assessment in New Zealand.

Moreover, we did not have an even number of participants in each condition. This would have been ideal, however, despite these limitations of our study demographic and different numbers in each condition, a major strength of our study design was the counterbalancing method we used for the conditions. Due to this counterbalancing design, we were able to eliminate any confounding variables that may have impacted our results.

Additionally, we also used the same assessor for both assessments for each participant. Therefore, when participants completed session 1 and session 2, they did so with the exact same assessor. This helped to ensure that there were minimal differences in administration techniques for each participant, which again eliminated any confounding variables that may have impacted our results.

We also did not calculate participants' IQ scores which prevented us from determining if our groups of participants were equivalent in terms of IQ, which would eliminate the chance that our findings were due to one group having a higher or lower IQ overall.

Our research has helped to fill a large gap in the literature around mask-wearing and teleneuropsychological assessment in the wake of COVID-19. We completed the first study looking specifically at memory tests and whether these changes in assessment conditions significantly impact people's working and short-term memory performance compared to standardised conditions. No other research currently in this literature has looked at cognitive domains and memory functioning specifically.

Recommendations for Future Research

We used a 2x2 mixed methods/repeated measures design. Where participants completed both sessions either on zoom or in person, and only the mask conditions were completed within subjects. One recommendation for future research would be to ensure that participants have a distraction-free environment when completing the assessment through

Zoom. It was difficult to ensure that there would be no distractions caused by uncontrollable variables e.g. outside noise, other people in the house etc. but a study by Mahon et al. (2022) requested participants move the camera around the room to ensure that there was no one else in the room during the assessment and the room was free of distractions. This may be a good idea for future research to make sure that the zoom conditions are as similar as possible to the controlled in-person conditions.

It is possible that participants answered the satisfaction questionnaire with evidence of confirmation bias. Responding in a way that favours what they believe is the assessor/researcher's opinions. So perhaps rather than asking participants how they found the session, allowing them to complete anonymous surveys may have helped to get a more accurate representation of their thoughts around the different assessment conditions. Moreover, participants did not have anything to compare their session to in regards to the visual stimulus, privacy, etc. scales so creating a within-subjects variable with Zoom and in person would allow participants to have a direct comparison and evaluate their opinions based on that. However, this may not be feasible as participants would be completing four different conditions which may cause fatigue.

It may be interesting to look at the effects of different types of masks on people's memory performance as we only used surgical masks because they are most commonly used. Previous research also suggests that surgical masks have the least impact on speech intelligibility, so perhaps other masks like N95 may impact memory performance even more by muffling speech sounds (Magee et al., 2020). Research does also suggest that people rate clear masks more positively than covered masks and have higher trust in healthcare professionals, while also helping communication and understanding (Kratzke et al., 2021).

Marler & Ditton (2021), suggested that the assessor should show the patient a picture of themselves without a mask on to help create a personal connection and recognition in future assessments. It would be interesting to see if this could help contribute to breaking down the cultural barrier that mask-wearing creates. The participant mentioned that in the second session she felt more comfortable as she got to know the assessor and now can see what she looked like (without a mask). Therefore, if that participant was shown a photo of that assessor before the first session began, that may help break down the cultural barrier and work towards building connection and whakawhanaungatanga as she would be able to view the assessors face which is important to Māori culture and feeling comfortable and at ease in a foreign environment. However, this could extend to other cultures too.

Conclusion

In conclusion, mask-wearing and remote teleneuropsychological assessments do not impact people's working memory and short-term memory on neuropsychological tests in a population aged 18-59 with no cognitive impairments. More specifically on Verbal Fluency, Digit Span and Verbal Learning subtests. There were no significant differences between people's performance when they completed the assessment on Zoom compared to in-person and when the assessor was wearing a mask compared to no mask. We did find an interaction effect for trial 5 of Verbal Learning, where participants who completed the assessment on zoom and with the assessor wearing a mask performed worse, with less words recalled. This is explained by the increased cognitive load pressure in this condition impacting the participants' working memory and short term memory. However, this effect was not clinically significant so not meaningful to the neuropsychological clinic space in New Zealand. Moreover, while not clinically significant, the finding that the alternate forms of RAVLT were not equivalent is a meaningful finding. This finding is important for neuropsychologists in New Zealand to be aware that when using alternate forms of the RAVLT the Crawford Verbal Learning test is not equivalent so proceed with caution. Overall, these are positive results to validate past neuropsychological assessment conditions during the COVID-19 pandemic, and future assessments when teleneuropsychological assessments or the use of face masks is required in a specific younger population.

References

Accident Compensation Corporation (n.d). *Injuries we cover*. https://www.acc.co.nz/im-injured/what-we-cover/injuries-we-cover/

Agrell, B., & Dehlin, O. (1998). The clock-drawing test. *Age and ageing*, *27*(3), 399-404.

Aguilar, M. J. G., & Leguizamón, L. D. (2021). Virtual versus face-to-face neuropsychological assessment: A pilot study towards its benefits and its limitations. *Global Journal of Aging & Geriatric Research, 1*(4), 1-4.

Alegret, M., Espinosa, A., Ortega, G., Pérez-Cordón, A., Sanabria, Á., Hernández, I., Marquié, M., Rosende-Roca, M., Mauleón, A., & Abdelnour, C. (2021). From face-to-face to home-to-home: Validity of a teleneuropsychological battery. *Journal of Alzheimer's Disease*, *81*(4), 1541-1553. https://doi.org/10.3233/JAD-201389

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child development*, *77*(6), 1698-1716. https://doi.org/10.1111/j.1467-8624.2006.00968.x

Baddley, A., Emslie, H., & Smith, I.N. (1992). The Speed and Capacity of Language-Processing Test. Pearson.

Baddeley, A. D. (2015). Working memory in second language learning. In Z, Wen., M.B, Mota., & A, McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 17-28). Multilingual Matters.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 570. https://doi.org/10.1037/0278-7393.33.3.570

Bean, J. (2011) 'Rey Auditory Verbal Learning Test, Rey AVLT'. In J.S. Kreutzer, J. DeLuca, 424 and B. Caplan (Eds.), *Encyclopedia of Clinical Neuropsychology* (pp. 2174-2175). Springer.

Bottalico, P., Murgia, S., Puglisi, G. E., Astolfi, A., & Kirk, K. I. (2020). Effect of masks on speech intelligibility in auralized classrooms. *The Journal of the Acoustical Society of America*, *148*(5), 2878-2884. https://doi.org/10.1121/10.0002450

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

Brown, V. A., Van Engen, K. J., & Peelle, J. E. (2021). Face mask type affects audiovisual speech intelligibility and subjective listening effort in young and older adults. *Cognitive Research: Principles and Implications*, *6*(1), 1-12. https://doi.org/10.1186/s41235-021-00314-0

Chapman, J. E., Cadilhac, D. A., Gardner, B., Ponsford, J., Bhalla, R., & Stolwyk, R. J. (2021). Comparing face-to- face and videoconference completion of the Montreal Cognitive Assessment (MoCA) in community-based survivors of stroke. *Journal of Telemedicine and Telecare, 27*(8), 484 – 492. https://doi.org/10.1177/1357633X19890788

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. Advance online publication. https://doi.org/10.1177/1357633X19890788

Clime, E.A., & Rostad, K. (2011). Test review: Wechsler Adult Intelligence Scale. *Journal of Psychoeducational Assessment, 29*(6), 581-586. https://doi.org/10.1177/0734282911408707

Cohen, R.J., & Swerdlik, M.E. (2005). *Psychological testing and assessment.* McGraw-Hill.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Crawford, J.R., Stewart, L.E., & Moore, J.W. (1989). Demonstration of savings on the AVLT and development of a parallel form. *Journal of Clinical and Experimental Neuropsychology, 11*(6), 975-981. https://doi.org/10.1080/01688638908400950

Cullum, C. M., Hynan, L., Grosh, M., Parikh, M., & Weiner, M. (2014). Teleneuropsychology: Evidence for video teleconference-based neuropsychological assessment. *Journal of the International Neuropsychological Society*, *20*(10), 1028-1033. https://doi.org/10.1017/S1355617714000873

Cullum, C. M., Weiner, M. F., Gehrmann, H. R., & Hynan, L. S. (2006). Feasibility of telecognitive assessment in dementia. *Assessment*, *13*(4), 385-390. https://doi.org/10.1177/1073191106289065

Darke, S. (1988). Anxiety and working memory capacity. *Cognition and Emotion*, *2*(2), 145-154. https://doi.org/10.1080/02699938808408071

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D–KEFS)*. Pearson. https://doi.org/10.1037/t15082-000

Dudley, M., Wilson, D., & Barker-Collo, S. (2014). Cultural invisibility: Māori people with traumatic brain injury and their experiences of neuropsychological assessments. *New Zealand Journal of Psychology*, *43*(3), 14-21.

Giovanelli, E., Valzolgher, C., Gessa, E., Todeschini, M., & Pavani, F. (2021). Unmasking the difficulty of listening to talkers with masks: lessons from the COVID-19 pandemic. *I-Perception, 12*(2), 1-11. https://doi.org/10.1177/2041669521998393.

Gnassounou, R., Defontaines, B., Denolle, S., Brun, S., Germain, R., Schwartz, D., Schück, S., Michon, A., Belin, C., & Maillet, D. (2022). Comparison of neuropsychological assessment by videoconference and face to face. *Journal of the International*

Neuropsychological Society, *28*(5), 483-493. https://doi.org/10.1017/S1355617721000679

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

Groth-Marnat, G., & Wright, J.A. (2016). *Handbook of psychological assessment.* John Wiley & Sons.

Gutz, S. E., Rowe, H. P., Tilton-Bolowsky, V. E., & Green, J. R. (2022). Speaking with a KN95 face mask: a within-subjects study on speaker adaptation and strategies to improve intelligibility. *Cognitive Research: Principles and Implications*, *7*(1), 1-20. https://doi.org/10.1186/s41235-022-00423-4

Harder, L., Hernandez, A., Hague, C., Neumann, J., McCreary, M., Cullum, C. M., & Greenberg, B. (2020). Home-based pediatric teleneuropsychology: A validation study. *Archives of Clinical Neuropsychology*, *35*(8), 1266-1275. https://doi.org/10.1093/arclin/acaa070

Harvey, P. D. (2022). Clinical applications of neuropsychological assessment. *Dialogues in clinical neuroscience*, *14*(1), 91-99. https://doi.org/10.31887/DCNS.2012.14.1/pharvey

Hess, D. C., & Audebert, H. J. (2013). The history and future of telestroke. *Nature Reviews Neurology, 9(*6), 340–350. https://doi.org/10.1038/nrneurol.2013.86

Kastendieck, T., Zillmer, S., & Hess, U. (2022). (Un)mask yourself! Effects of face masks on facial mimicry and emotion perception during the COVID-19 pandemic. *Cognition and Emotion*, *36*(1), 59-69. https://doi.org/10.1080/02699931.2021.1950639

Kirkwood, 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*(3), 147-151.

Kratzke, I. M., Rosenbaum, M. E., Cox, C., Ollila, D. W., & Kapadia, M. R. (2021). Effect of clear vs standard covered masks on communication with patients during surgical clinic encounters: a randomized clinical trial. *JAMA Surgery*, *156*(4), 372-378. https://doi.org/10.1001/jamasurg.2021.0836

Laatsch, L. (2002). Neuropsychological Assessment. In M. Hersen & W. Sledge (Eds.), *Encyclopedia of Psychotherapy* (2nd ed., pp. 223 – 228). Academic Press.

Lee, E., Cormier, K., & Sharma, A. (2022). Face mask use in healthcare settings: effects on communication, cognition, listening effort and strategies for amelioration. *Cognitive Research: Principles and Implications*, *7*(1), 1-9. https://doi.org/10.1186/s41235-021-00353-7

Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological assessment* (4th ed.). Oxford university press.

Lichtenstein, J. D., Amato, J. T., Holding, E. Z., Grodner, K. D., Pollock, E. N., Marschall, K. P., & Scull, S. (2022). How we work now: Preliminary review of a pediatric neuropsychology hybrid model in the era of COVID-19 and beyond. *Archives of Clinical Neuropsychology*, *37*(1), 40-49. https://doi.org/10.1093/arclin/acab041

Magee, M., Lewis, C., Noffs, G., Reece, H., Chan, J. C., Zaga, C. J., Paynter, C., Birchall, O., Rojas Azocar, S., & Ediriweera, A. (2020). Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols. *The Journal of the Acoustical Society of America*, *148*(6), 3562-3568. https://doi.org/10.1121/10.0002873

Mahon, S., Webb, J., Snell, D., & Theadom, A. (2021). Feasibility of administering the WAIS-IV using a home-based telehealth videoconferencing model. *The Clinical Neuropsychologist*, *26*(3), 1-13. https://doi.org/10.1080/13854046.2021.1985172

Marler, H., & Ditton, A. (2021). "I'm smiling back at you": exploring the impact of mask wearing on communication in healthcare. *International Journal of Language & Communication Disorders*, *56*(1), 205-214. https://doi.org/10.1111/1460-6984.12578

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

Molloy, D. W., & Standish, T. I. (1997). A guide to the standardized Mini-Mental State Examination. *International psychogeriatrics*, *9*(S1), 87-94. https://doi.org/10.1017/S1041610297004754

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. https://doi.org/10.1258/ 1357633971931048.

Nesher Shoshan, H., & Wehrt, W. (2022). Understanding "Zoom fatigue": A mixed-method approach. *Applied Psychology*, *71*(3), 827-852. https://doi.org/10.1111/apps.12360

New Zealand Tourism (n.d). *Tā moko: Traditional Māori tattoo.* https://www.newzealand.com/nz/feature/ta-moko-maori-tattoo/

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. https://doi.org/10.1093/arclin/acab002

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. https://doi.org/10.1093/arclin/acab002

Rochette, A., Rahman-Filipiak, A., Spencer, R.J., Mashal, D., & Stelmokas, J.E. (2021). Teleneuropsychology practice survey during COVID-19 within the United States.

Applied Neuropsychology: Adult, 29(6), 1312-1322. https://doi.org/10.1080/23279095.2021.1872576

Roth, C. (2011). Boston Naming Test. In J.S, Kreutzer., J, DeLuca., B, Caplan (Eds), *Encyclopedia of Clinical Neuropsychology* (2nd ed., pp. 430-433). Springer. https://doi.org/10.1007/978-0-387-79948-3_869

Saint, S. A., & Moscovitch, D. A. (2021). Effects of mask-wearing on social anxiety: an exploratory review. *Anxiety, Stress, & Coping*, *34*(5), 487-502. https://doi.org/10.1080/10615806.2021.1929936

Saunders, G. H., Jackson, I. R., & Visram, A. S. (2021). Impacts of face coverings on communication: an indirect impact of COVID-19. *International Journal of Audiology*, *60*(7), 495-506. https://doi.org/10.1080/14992027.2020.1851401

Schmidt, M. (1996). *Rey Auditory Verbal Learning Test: A handbook*. Los Angeles, CA: Western Psychological Services.

Schnabel, R. (2012). Overcoming the challenge of re-assessing logical memory. *The Clinical Neuropsychologist*, *26*(1), 102-115. https://doi.org/https://doi.org/10.1080/13854046.2011.640642

Semrud-Clikeman, M. E., & Swaiman, K. F. (2017). Neuropsychological Assessment. In *Swaiman's Pediatric Neurology: Principles and Practice: Sixth Edition* (pp. 65-72). Elsevier Inc. https://doi.org/10.1016/B978-0-323-37101-8.00010-2

Sharma, M., Weerawardhena, H., Wall, A. K., VanderZwaag, B., Andruchow, D. L., Al-Khaz'Aly, H., Cunningham, D. R., & Callahan, B. L. (2021). Teleneuropsychological Assessment During the COVID-19 Pandemic: Where do we go from here? *Journal of Student Research*, *10*(2). https://doi.org/10.47611/jsr.v10i1.1178

Smith, C. J., Rozga, A., Matthews, N., Oberleitner, R., Nazneen, N., & Abowd, G. (2017). Investigating the accuracy of a novel telehealth diagnostic approach for autism spectrum disorder. *Psychological Assessment, 29*(3), 245–252. https://doi.org/10.1037/pas0000317

Spataro, J. (2020, July 8). *The future of work – the good, the challenging & the unknown.* Microsoft. https://www.microsoft.com/en-us/microsoft-365/blog/2020/07/08/future-work-good-challenging-unknown/

Stead, A., & Vinson, M. (2019). Cognitive assessment using face-to-face and videoconferencing methods. *Nursing Older People*, *31*(5), 34-39. https://doi.org/10.7748/nop.2019.e1160

Svenn, E. J., Terje, S., Stein, A., & Jan-Magne, K. (2003). Neuropsychological assessment and telemedicine: A preliminary study examining the reliability of neuropsychology services performed via telecommunication. *Journal of the International Neuropsychological Society 9*(3)*,* 472-478. https://doi.org/10.10170S1355617703930128

Tailby, C., Collins, A. J., Vaughan, D. N., Abbott, D. F., O'Shea, M., Helmstaedter, C., & Jackson, G. D. (2020). Teleneuropsychology in the time of COVID-19: The experience of the Australian epilepsy project. *Seizure*, *83*, 89-97. https://doi.org/https://doi.org/10.1016/j.seizure.2020.10.005

Thorndike, E. L., & Lorge, I. (1944). *The teacher's wordbook of 30,000 words*. New York: Teachers College, Columbia University.

Truong, T. L., Beck, S. D., & Weber, A. (2021). The impact of face masks on the recall of spoken sentences. *The Journal of the Acoustical Society of America*, *149*(1), 142-144. https://doi.org/10.1121/10.0002951

Turner, J. (2020, December 30). What are zoom calls doing to your brain? *Medium. https://medium.com/swlh/what-are-zoom-calls-doing-to-your-brain-ebe288d12806*

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. https://doi.org/10.1016/j.jagp.2014.08.006.

Wadsworth, H. E., Dhima, K., Womack, K. B., Hart Jr, 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. https://doi.org/10.1093/arclin/acx140

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

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale Fourth Edition (WAIS-IV)*. Pearson.

Wechsler, D. (2009). *Wechsler Memory Scale Fourth Edition (WMS-IV)*. Technical and Interpretive Manual. Pearson.

Wechsler D. (2011). *Wechsler Abbreviated Scale of Intelligence Second Edition (WASI-II).* Manual. Pearson.

Yi, H., Pingsterhaus, A., & Song, W. (2021). Effects of wearing face masks while using different speaking styles in noise on speech intelligibility during the COVID-19 pandemic. *Frontiers in Psychology*, *12*, 682677. https://doi.org/https://doi.org/10.3389/fpsyg.2021.682677

Appendices

Appendix A – Participation Flyer (Students)

The effects of remote test administration and face masks on neuropsychological test performance

What is this study about?

- Neuropsychological tests are used to assess attention, concentration, learning and memory. They are often used to assess the effects of a stroke or brain injury.
- The assessments are usually administered face-to-face in a quiet room but because of the COVID19 pandemic the tests may be administered remotely and/or the assessor has to wear a mask.
- We want to find out the effects of face masks on peoples' scores on some commonly used neuropsychological tests.
- The study is being led by Professor Nicola Starkey, School of Psychology, University of Waikato and has received approval from the University of Waikato Human Research Ethics Committee (HREC(Health)2022#04); humanethics@waikato.ac.nz)

Am I eligible to take part?

- You are aged between 18-80 years
- Your primary language is English
- Have not had a head injury, stroke or other medical condition that could affect cognitive function

What am I being asked to do?

- To complete some neuropsychological tests to assess your cognitive function either online or in-person. This will take about 1.5 hours. We would like you to complete the tests twice (with the assessor wearing a mask and without) approximately 2-weeks apart.
- To answer some questions about your experience of the assessment sessions
- To thank you for your participation you will be offered 5% course credit for each session

Who can I speak with about my participation in this project?

- Contact Nicola Starkey on 07 8379230 or email: nstarkey@waikato.ac.nz

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Appendix B – Participation Flyer (General Public)

The effects of remote test administration and face masks on neuropsychological test performance

What is this study about?

- Neuropsychological tests are used to assess attention, concentration, learning and memory. They are often used to assess the effects of a stroke or brain injury.
- The assessments are usually administered face-to-face in a quiet room but because of the COVID19 pandemic the tests may be administered remotely and/or the assessor has to wear a mask.
- We want to find out the effects of face masks on peoples' scores on some commonly used neuropsychological tests.
- The study is being led by Professor Nicola Starkey, School of Psychology, University of Waikato and has received approval from the University of Waikato Human Research Ethics Committee (HREC(Health)2022#04); humanethics@waikato.ac.nz)

Am I eligible to take part?

- You are aged between 18-80 years
- Your primary language is English
- Have not had a head injury, stroke or other medical condition that could affect cognitive function

What am I being asked to do?

- To complete some neuropsychological tests to assess your cognitive function either online or in-person. This will take about 1.5 hours. We would like you to complete the tests twice (with the assessor wearing a mask and without) approximately 2-weeks apart.
- To answer some questions about your experience of the assessment sessions
- To thank you for your participation you will be offered a $20 voucher for each session

Who can I speak with about my participation in this project?

- Contact Nicola Starkey on 07 8379230 or email: nstarkey@waikato.ac.nz

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Appendix C - Screening Questionnaire

PPT ID

The effects of remote test administration and face masks on neuropsychological test performance

Participant ID:____________________ **Date:** _____________**Time**: _________________

Researcher Name : _____________

Screening questionnaire

Certain illnesses and medication may impact your performance on some of the neuropsychological tests. So, the next few questions ask for some information about your medical history.

Are you currently taking any Medication? 1. No. 2. Yes: -

If yes, does it affect your cognitive ability (eg makes it hard to concentrate, makes you drowsy or are you advised not to drive after taking it?)

1. No. 2. Yes: -

Current Medical Conditions: [ ]
(Do you currently have any condition which affects your eyesight, hearing, sense of smell or touch, or conditions of epilepsy, diabetes, etc?)
1. No 2. Yes – Type: ____________________

Current Psychiatric Treatment: [ ]
(Are you currently being treated by a Psychiatrist or Psychologist?).
1. No 2. Yes – Diagnosis: _________________

Previous medical conditions
Have you had a stroke or been diagnosed with another neurological condition?
1. No 2. Yes – Diagnosis: _________________

Have you experienced a traumatic brain injury where you were unconscious for more than 20 minutes?
1. No 2. Yes – When: _________________

Do you have any other health conditions that might affect your performance on the tests?

1. No 2. Yes – Details: _________________

If participant answers yes to any of the above questions they are not eligible for the study

Appendix D – Participant Information Sheet

The effects of remote test administration and face masks on neuropsychological test performance

Participant Information Sheet

What is this study about?

When someone has an injury or illness that might affect their brain (e.g. brain injury or stroke) assessments are carried out to find out what effect this may have on their day to day lives and also to monitor recovery. The tests we use (neuropsychological tests), have been designed to be administered under specific testing conditions, usually face-to-face with the assessor in a quiet room free from distractions. The COVID-19 pandemic has meant that these tests can no longer be administered under the recommended testing conditions as we need to limit the spread of COVID. So, tests may be administered remotely and/or for in person tests the assessor may have to wear personal protective equipment (PPE) including masks.

We are interested in finding out how these changes to test administration affect peoples' performance on these tests. This will help us to provide information to psychologists to help ensure test scores are accurate.

The study is being led by Professor Nicola Starkey, School of Psychology, University of Waikato with the assistance of two Masters students, Tiffany Rich and Olivia Benge, and two Honours students, Daniel Mackie and Maddison Beaumont. Dr Shirley Hosking is the neuropsychology advisor.

Am I eligible to take part?

You are eligible to take part in this study if you are aged 18 – 80 years of age, your primary language in English, and have no history of stroke, head injury or other medical problems that impacts your cognitive function (including medication).

What am I being asked to do?

If you agree to take part in this study, it will involve two sessions of approximately 1.5 hours. You will be asked to take complete a series of cognitive assessments, which most people find enjoyable. We will also ask for some information about you and also ask you to tell us about your experience of the assessment. As an acknowledgement of your contribution to the research we will offer you a $20 voucher for each assessment. You can complete the assessments online or at the University (depending on your preference and the COVID restrictions in place at the time). Your participation is voluntary (your choice).

What will happen to my information?

Once we have scored the assessments, the information you provide will be anonymised. All paper forms will be stored in a locked cabinet, in the School of Psychology at Waikato University and electronic data will be stored in a password protected folder. At the end of the study the paper-based forms will be destroyed. We will send an electronic summary of our findings to the participants who have indicated they would like to receive this information. The study findings will be written up for publication as a journal article, and included as two Masters theses and two Honours theses.

Appendix D – Participant Information Sheet Continued

If the scores you obtain on the tests are unusual we will consult with the neuropsychology advisor for the study (Dr Shirley Hosking), contact you to let you know and provide you with further advice.

What are the possible benefits and risks of this study?

Taking part in this study will take some of your time. There are no known risk caused by this study, however you may feel uncomfortable about answering some of the questions. You do not have to answer any questions you do not wish to do so.

This study will be of benefit to the wider population. There is no guarantee that you will benefit directly from being involved in this study. The results obtained from your participation will help us understand how changing test conditions might affect peoples' scores on neuropsychological tests.

What can I expect from the researchers?

If you decide to participate in this project, the researchers will respect your right to:

- ask any questions of the researchers about the study at any time during participation; • decline to answer any particular questions or carry out any of the tasks;
- withdraw from the study up to one week after completing the assessment;
- provide information on the understanding that it is confidential to the researchers. All forms are identified by a code number, and are only seen by the researchers. It will not be possible to identify you in any articles produced from the study;
- be given an electronic summary of the findings

Who can I speak with about my participation in this project?

If you, or anyone you know is interested in taking part in this research please contact Professor Nicola Starkey on 07 8379230 or email nstarkey@waikato.ac.nz for further information.

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Appendix E – Consent Form

The effects of remote administration and masks on neuropsychological test performance

Consent Form

Please tick to indicate you consent to the following Yes

I have read, or have had read to me in my first language, and I understand the Participant Information Sheet. □

I have been given sufficient time to consider whether or not to participate in this study. □

I am satisfied with the answers I have been given regarding the study and I have a copy of this consent form and information sheet. □

I understand that taking part in this study is voluntary (my choice) and that I may withdraw from the study within a week of the assessment □

I understand that my participation in this study is confidential and that no material, which could identify me personally, will be used in any reports on this study. □

I know who to contact if I have any questions about the study in general. □

I understand my responsibilities as a study participant. □

I wish to receive a summary of the results from the study. Yes □ No □
If yes, please provide your name and email address so we can send a summary of the results

Declaration by participant:
I hereby consent to take part in this study.

Participant's name:

Signature: Date:

Declaration by member of research team:

I have given a verbal explanation of the research project to the participant, and have answered the participant's questions about it. I believe that the participant understands the study and has given informed consent to participate.

Researcher's name:

Signature: Date:

Appendix F – Demographic Questionnaire

The effects of remote test administration and face masks on neuropsychological test performance

Participant ID:____________________ **Date:** _____________**Time**: _________________
Researcher Name : _____________

Demographic questions (session 1 only)

Age: ____________________________

What gender do you most identify as?

- Male
- Female
- Non-binary/third gender
- Prefer to self-describe: ______________________
- Prefer not to say

What is your ethnicity? (select as many as apply)

New Zealand European
Māori
Samoan
Cook Island Māori
Tongan
Niuean
Chinese
Indian
Other such as Dutch, Japanese, Tokelauan (please specify)

First language: _________________________

Education:
Years of secondary Education: ______________________________________
Years of tertiary Education: ____
Age when left school: ____
Years of School Education: ____

Occupational Training Qualifications:
1. Nil
2. School Certificate
3.U.E./Sixth Form Certificate.
4. HSC/Bursary.
5. Trade – Specify. ____
6. Professional – Specify.____
7. Tertiary – Specify.____
8. Other – Specify.____

Appendix G – Satisfaction Questionnaire

PPT ID:

Satisfaction with the test session (session 1 and 2) – circle session number

This section asks about your experience of the test session.

1. Have you used video conferencing software before (e.g. zoom)? Yes No

Please rate the following aspects of the session:

	Very dissatisfied	Somewhat dissatisfied	Neither	Somewhat satisfactory	Very satisfactory
2. Audio quality	1	2	3	4	5
3. Visual stimulus quality	1	2	3	4	5
4. Privacy	1	2	3	4	5
5. Comfort	1	2	3	4	5
6. Convenience	1	2	3	4	5
	Not at all confident	Somewhat confident	Moderately confident	Very confident	Completely confident
7. Confidence using computers	1	2	3	4	5

8. Did you perceive any cultural barriers during the test session?

9. Please describe any difficulties or barriers experienced during the test session

10. Is there any other feedback you would like to provide about the test session?

End of session

Thank participant for their time and answer any questions they might have.