DPhil thesis: Post-stroke neuropsychological screening and assessment: psychometrics, validation, and predictive value for functional outcomes



Sam S. Webb

Student number - 1203297

University of Oxford

Supervisors - Prof. Nele Demeyere & Prof. Sarah Pendlebury

Viva examiners - Prof. Jonathan Evans & Dr. Antoniades

Acknowledgements

"You better work." - Supermodel (You better work) by RuPaul.

I would like to begin by thanking my supervisor, Prof. Nele Demeyere, for her exceptional supervision of my research starting from my MSc in 2018, to the completion of my DPhil in 2024. Years ago, I was told the two most important decisions in life were 1) who I married, and 2) who my PhD supervisor was. I could not be more grateful for ensuring I made the right choice of supervisor. On meeting Nele at the MSc induction event in 2018, I knew immediately that I wanted to work with her, whatever she did for research, I would do it. Years later, I still wake up every day and look forward to conducting research within her team and am excited about future projects. Not many people can say that honestly, and I do, wholeheartedly. Note, I do have a life outside of work, I swear. So I thank you, Professor Nele Demeyere, for your endless support, motivation, encouragement, banter, honesty, and for being a worthy role model and mentor.

I also want to thank my secondary supervisor Professor Sarah Pendlebury and collaborators for their efforts and support for my research. I especially thank Sarah for her medical and clinical perspective, always ensuring my work is meaningful for practice not just for science. My lab mates over the years have made slogging through data collection and daily life the best experience. No day passes in the office without laughs, inside jokes, banter, support, friendship, and compassion.

Next, I want to thank my DPhil examiners, including those who have helped with my Transfer of Status, Confirmation of Status, and DPhil viva. It is with these crucial steps in my DPhil, that I have grown my research, my framework for this thesis, and have been given the opportunity to think about my work critically, deeply, and with humility.

The Stroke Association took a chance by funding me and my project, and I am eternally grateful that they saw potential and spearheaded what I hope will be a long research career in stroke. Particular thanks are given to the communication team, including Clare Jonas who always supported me and gave me so many opportunities to grow and disseminate my research.

I want to thank the Oxford Health NHS Trust and Oxfordshire University Hospitals NHS trust. Both provided unparalleled access to recruitment opportunities

not seen often in the UK. Particularly, Oxford Health NHS Trust allowed me to work within the Oxfordshire Stroke Rehabilitation Ward (OSRU) with unlimited access and full support. It is with the support of occupational therapists and physiotherapists at OSRU that I have succeeded. Particular thanks are given to Sarah King, Vivienne Southcott, and Michelle Fellerdale, and Emma Garrett. You're all amazing! They each gave me constant critical feedback on my projects, enabling me to always consider their perspective on what is actually feasible, what is actually clinically relevant, and what is actually needed. The team at OSRU helped me mentally, being the only team I had physical access to during 2020's lockdowns and restrictions. It was being able to come to OSRU, fully kitted out in protective equipment to do my research, and see friendly faces (half of them due to the masks) and have fun conversations, that kept me going.

Crucially, my research is only worthwhile because of the willingness of so many individuals to take part. I thank the stroke survivors who I met on the stroke rehab unit, who gave me their time and their mental efforts shortly after a life changing event. I appreciate, deeply, that within the chaos of having had a stroke and treatments and transfers and so much uncertainty, that they trusted me to help me with my research. Perhaps for them it was just to pass the time, but for me their contributions made my research happen. I also thank the many healthy ageing adults who took part, many of whom were loved ones of the stroke survivors. For many, taking part in my research was an insight into what their loved one was going through. For others, it just seemed interesting after seeing our adverts for participants. For me, again, without their volunteering I could not have completed my research and more projects in the future.

Finally, I must thank my family and friends who have support me throughout the last four years. For most people a PhD, or DPhil, means nothing, it is an abstract term used by nerds and boffins. As a first generation university student, and certainly the first to go onto postgraduate education, my family were behind me 100%. My family helped fund my degrees, housed me, clothed me, and fed me. I thank my family so much for their support. My friends, both existing and new since coming to Oxford, have also always supported me and tried to understand my work. Special thanks are given to Leonie Glitz who provided ample hugs and emotional support for me over the years, and I hope I can have even half the compassion she exudes. Even though my experimental psychology friends describe my research as 'just running some correlations', they appreciate the impact those 'correlations' can have, and the benefit they bring to stroke survivors. To my friends and loved ones, I thank you.

Thank you all.

Abstract

Neuropsychological screening and assessment of cognitive changes following stroke requires the use of standardised and well validated tools to ensure the most accurate and evidence-based assessment of a patient's abilities.The importance of cognitive assessment comes from the high relation between cognitive impairment and later recovery [1–3]. In my DPhil, I aimed to examine different cognitive screening tools and provide validity evidence for their use in assessing post-stroke cognitive impairment.

The first chapter examined the shift in modality of cognitive assessment during the COVID-19 pandemic in a completed web-based questionnaire of 114 allied health professionals (majority UK). This study was conducted at the height of the pandemic and provided time sensitive information about how the use of cognitive testing had changed. I found that the frequency of cognitive testing did not necessarily decrease, but that the modality changed to more remote and video based assessments. The rates of change were lower than other studies at the time, that primarily sampled neuropsychologists from the United States. Within the same chapter, I detailed a study, that provided some evidence for the valid use of a remote version of the Oxford Cognitive Screen (OCS). This study used 40 stroke survivors who had been administered the OCS both in person and remotely, in counterbalanced order. I found that there were minimal meaningful differences in performance across versions, and limited occurrences of a participant being considered impaired on one version and not on the other. This meant that, in light of the pandemic, I had provided some validity evidence for a tool usable via remote administration.

The second chapter provided some clinical validation evidence of a computer tablet-based extension of the OCS, named the OCS-Plus. OCS-plus is a recently standardized and normed domain-general cognitive screening tool. Prior to this DPhil, there was only psychometric validation in 320 healthy ageing adults and no evidence for use in stroke. Here I examined OCS-Plus data from 347 stroke survivors (n=181 <3 months post stroke & n=166 >=6 months post-stroke) and compared performance of a subset of them to a neuropsychological battery and/or the Montreal Cognitive Assessment (MoCA). I found that the OCS-Plus met convergent and divergent validation criteria compared to the neuropsychological battery, and detected cognitive impairments in a large proportion of those classed

as unimpaired on MoCA (100%) and OCS(98.50%). This project showed that the OCS-Plus is valid as a screening tool in stroke from subacute to chronic usage.

The third chapter detailed work regarding the Oxford Digital Multiple Errands Test (OxMET), which had previously been standardised and normed in 124 healthy controls and compared to 105 chronic (>6 months post-stroke) stroke survivors. In part one of chapter 3, I examined the OxMET in relation to the in-person Multiple Errands Test - Home version (MET-Home) in 98 participants, to determine evidence for convergence between the MET, as well as feasibility and acceptability for each MET. I recruited 48 stroke survivors and 50 healthy ageing adults to take part. I found that both METs met convergence criteria (r>=.30). I found the OxMET to be more feasible and acceptable by participants, but that there was an appropriate time and sample to use for each MET and thus both are satisfactory tools.
In part 2 of chapter 3, I examined the predictive validity of the OxMET taken subacutely with stroke survivors to clinically relevant future functional outcomes. The findings were clear that the OxMET was associated with both basic (n=117) and instrumental (n=66) activities of daily living, although more so with instrumental, over and above the contribution of other demographic and cognitive factors.

The aim of the work presented in this thesis was to provide clinicians with an evidence base to make judgement about which tools are best practice for use, and to highlight both strengths and weaknesses of each tool presented. None of the tests presented in this thesis are able to fully capture the full depth of human cognition nor the rich idiosyncrasies of real everyday behavior. Health care professionals should always follow latest guidance and be educated in best evidence-based screening use and trained in tools they administer.

Contents

List of Abbreviations **viii**

Introduction **1**

0.1 Establishing psychometric properties of neuropsychological assessments 2
0.2 Assessment of primary/gross level cognitive impairments post-stroke 4
0.3 Assessment of complex and/or subtle cognitive impairments post-stroke 13
0.4 Ecological validity of assessments of cognitive impairments post-stroke 17
0.5 Thesis structure 25

1 Remote cognitive screening **28**

1.1 The COVID-19 pandemic altered the modality, but not the frequency, of formal cognitive assessment 28
1.2 Introducing the Tele-OCS: A validated remotely administered version of The Oxford Cognitive Screen 47

2 Screening for subtle cognitive deficits **61**

2.1 Validation of the UK English Oxford Cognitive Screen-Plus in Sub-Acute and Chronic Stroke Survivors 61

3 Ecological and predictive validity of Multiple Errands Tests **85**

3.1 Comparing the Oxford Digital Multiple Errands Test (OxMET) to a real-life version: convergence, feasibility, and acceptability. 85
3.2 Predictive validity of the Oxford Digital Multiple Errands Test (OxMET) for functional outcomes after stroke. 112

Discussion **126**

3.3 Assessment of primary/gross level cognitive impairments post-stroke 127
3.4 Assessment of complex and/or subtle cognitive impairments post-stroke 130
3.5 Ecological validity of assessments of cognitive impairments post-stroke 133
3.6 Validity of neuropsychological tests and limitations in the presented works 136
3.7 Integrating validity evidence: directions for future research 143

Conclusion **148**

4 Appendix A **150**

4.1 OCS 150

4.2 OxMET 156

4.3 OCS-Plus 156

5 Appendix B **162**

6 Appendix C **167**

7 Appendix D **172**

7.1 New normative data and concurrent validity for the MET-Home revised in english speaking healthy adults and stroke survivors . . . 172

7.2 Tables and Futher data 193

References **205**

List of Abbreviations

COVID-19 . . Novel corona virus disease 2019

OCS . . . . . Oxford Cognitive Screen

OCS-Plus . . . Oxford Cognitive Screen - Plus

OxMET . . . . Oxford digital Multiple Errands Test

MoCA . . . . . Montreal Cognitive Assessment

MMSE . . . . Mini-Mental State Examination

BADS . . . . . Behavioural Assessment of Dysexecutive Syndrome

Introduction

Stroke, which is a disruption of blood flow to the brain either as a bleed or blockage/infarction of blood vessels [4], is a world leading cause of disability [5] and death [4]. With improved medical intervention and prevention, the incidence of stroke is declining [6] and the rate of survival from stroke is increasing [7]. Nearly all stroke survivors have some form of cognitive impairment within a few days or weeks of stroke, and many cognitive impairments are still evident 6-months post-stroke and beyond [8–13]. Even five years post-stroke, over 80% of stroke survivors report experiencing multiple cognitive problems [14]. Cognitive impairment has been shown to impact activities of daily living over and above the effects of physical impairments and stroke severity [15–18]. Failure to detect impairments in cognition can lead to failure to target rehabilitation to address or attenuate these impairments, and can increase the severity of disability and incidence of mortality [19]. As disabilities in motor functions and cognition are common following stroke, and as survival is increasing, more people are living with stroke-related impairments than ever before. This is likely to result in increased demands on health and social services [20]. Stroke remains an important and disabling condition and the need for immediate medical intervention as well as chronic care post-stroke are paramount. Particular care is needed in the diagnosis and assessment of cognitive impairments to attenuate disability and reduce incidence of post-stroke mortality. In order to directly assess and diagnose cognitive impairments, valid and reliable neuropsychological assessments must be used. Currently, there are few stroke-specific valid, reliable, and accessible measures. This thesis presents multiple completed works assessing the reliability, validity, and acceptability of multiple cognitive assessments, to better clinical practice, and benefit the lives of stroke survivors.

In this chapter, I will first detail how researchers establish the psychometric properties of neuropsychological assessments which lays the ground work for all measures and chapters discussed in this thesis. Second, I will describe and discuss how clinicians and researchers assess cognition on a primary/gross level to highlight clear impairments in specific cognitive domains. Third, I will detail the impact of the COVID-19 pandemic on assessment of cognition. Fourth, I will describe assessment of more subtle impairments in cognition which are often missed on primary/gross level screening assessments. Finally, I will change perspective to assessment of cognition in a way that reflects real life performance.

0.1 Establishing psychometric properties of neuropsychological assessments

First, this section describes validity as a construct and how it applies to neuropsychological assessment in terms of the research that underpins tests used in clinical practice. When diagnosing and screening for cognitive impairments post-stroke, in order to deduce secondary effects of stroke, validated neuropsychological assessments should be used. The validity of neuropsychological assessments depends on the validity evidence generated; whether it does measure the construct it is supposed to, whether it was normed on relevant healthy and clinical samples where appropriate; whether it is being used in the intended context. When validating measures, researchers must consider the standards of validation [21], and fully justify all choices that lead to whether a test is considered validated or not. Following the Standards, this thesis largely focuses on validating constructs, validation of tools on intended respondent populations, and further providing validity evidence in terms of predictive and convergent validity. As such, this section describes key concepts and ideas for validity evidence.

Test validity is a multifaceted construct. It encompasses not only the meaning of test-score interpretation, but also the context of test administration, the person taking the test, and the theoretical constructs being adequately measured by the test

[22]. Validation refers to the process of establishing evidence for the interpretation of tests used in their intended populations/settings. Validation is a never-ending process of generating evidence for the proper use of a tool [23]. The 2014 Standards for Educational and Psychological Testing (standards 1.12 & 1.16) state that if a cognitive process is being assessed via a test, then evidence must be generated that the test does assess that cognitive process [21], and to generate evidence using well construct-validated tools. Best practice for a full validation requires multiple projects examining different aspects of the validity of the test and test interpretation [24].

A 2013 systematic review of validation procedures for neuropsychological batteries aimed at clinical populations, examined 147 studies that were published between 2005 and 2012 [25]. The clinical populations in the reviewed studies predominantly included dementia and Alzheimer's disease, mild cognitive impairment, and acquired brain injury, though other clinical groups were also present. In terms of sample group used, the systematic review found that half of studies included both healthy controls and a clinical group, followed by 35.40% that included just clinical groups, and 13.50% that included just healthy controls [25]. Sample sizes ranged from 13 to 705 people for clinical groups and three to 13,000 healthy controls, with the majority of papers having fewer than 100 participants per group [25]. The sample group and size should be pre-defined depending on the intended target group for the test (e.g., if the test is developed for diagnosing dementia, it is appropriate to validate the test on those with dementia), and whether it is used to determine impairment compared to normative cut offs (e.g., generating validity evidence from healthy control participants, and using the normative cut offs to diagnose impairment in a clinical group) [21]. Sample size is often arbitrary in clinical research with no clear a priori justification [26], and sample size justifications differ dependent on the type of validity evidence being collected (e.g., factor analysis to investigate internal factor structure of a scale versus regressions for predicting outcomes) [27]. Sample size impacts validation outcomes because the power to detect effects will be drastically reduced the smaller the sample is [28, 29], and additionally small samples could reveal a large correlation given a less reliable estimate of the correlation [30].

When examining whether or not a test measures a construct, or cognitive process, it is important to acknowledge a priori which test metric is used and for what purpose, as is standard [21]. In practice, authors do not often justify the choice of a metric explicitly, by discussing either the psychometrics of the score or the intended use of the metric [31]. Typically, researchers refer to previous studies or clinical practice, although research may mention that there are alternative metrics (see example Arbuthnott and Frank [32]). Metric justification is particularly important where different metrics differentially relate to assessments of the same construct [33]. Further, norms for neuropsychological assessments are often stratified by age and education [34] due to the known influence of these variables on cognitive performance. Similarly, the inclusion of known covariates is important and common in neuropsychological assessments [35] (cf Lowe and Reynolds [36]). Importantly, not including age or education in neuropsychological test validation analyses can incorrectly inflate the correlation coefficients observed as they are both related to performance on the tests. In other words, the inclusion of known covariates will appropriately attenuate correlations, impacting validation outcomes.

In summary, each aspect of a given study is important to justify when generating validity evidence so researchers and clinicians know the strengths and limitations of the supporting evidence behind tools they use.

0.2 Assessment of primary/gross level cognitive impairments post-stroke

In this section, I will describe cognitive impairment screening of primary/gross level cognitive domains, and I will contrast and compare commonly used cognitive screening tools. Primarily discussion will focus on the The Montreal Cognitive Assessment (MoCA) [37] and the Oxford Cognitive Screen (OCS; [38]) as one of the historically most used screens and the newest established and validated stroke-specific screen, respectively.

There is a distinction between *domain-general* and *domain-specific* cognition, whereby domain-specific refers to a domain of cognition such 'language' and domain-general refers to cognition as a whole, which is likely to impact on domain-specific cognitive abilities. These are typically larger "umbrella" cognitive abilities such as executive function and memory. In memory for example verbal language abilities and verbal memory abilities group together; see Demeyere et al. [39]. In current stroke guidelines, cognitive testing is recommended in acute settings shortly after a stroke [40–42], and the cognitive domains instructed to assess include: attention, visual perception, memory, and executive functioning and apraxia [41, 43, 44]. There is, however, limited to no consensus on how to go about cognitive testing, what tests to use, when to conduct the testing, and which exact domains to assess, either specific or in general [44]. In a 2021 systematic review of clinical practice guidelines, it was found that out of eight eligible guidelines, only one specified a cognitive assessment tool, and the rest recommended 'validated' assessment tools but with no specific suggestions [44]. The lack of recommendation leaves the choice of tool up to the stroke clinicians who are not normally trained in cognitive assessment yet are expected to perform such tests [45]. Specific recommendations are excluded from guidelines to ensure that the guidance stays relevant through advances in test development. This is not useful for clinicians however who seek guidance.

The most recent stroke guidelines require that only clinicians trained in administration and interpretation of a specific tool and who are familiar with its limitations may conduct cognitive assessments [43], this is simply not always done in practice. Furthermore, only 7% of UK stroke centers have access to a clinical psychologist who is trained in assessment of mood and cognition [44]. While there is a shortage of trained professionals in cognitive assessment, UK specific stroke guidelines require assessments of cognition and mood post-stroke.The presence of cognitive screening is audited in all UK stroke centres as part of The Sentinel Stroke National Audit Programme (SSNAP) based upon the National Clinical Guideline for Stroke [43, 46]. The SSNAP only requires cognitive screening, rather than a full neuropsychological

assessment of domain-specific cognitive abilities, but the national stroke guidelines do require comprehensive testing in specified circumstances [43].

A cognitive *screen* refers to a brief assessment of cognition either on a domain-general or domain-specific level to indicate possible cognitive impairment [47]. If cognitive impairment is found, the stroke survivor may then be administered a comprehensive neuropsychological battery, however, given the lack of clinical psychologists this is often not the case [44]. Cognitive screens come in many shapes and forms. The Montreal Cognitive Assessment (MoCA) [37] and the Mini-Mental State Examination (MMSE) [48] are the most commonly used paper and pen based tools to screen for global cognitive impairment following stroke [45, 47, 49] because they are quick to administer and produce an easy to interpret cut off score [50]. Another commonly used cognitive screening tool is the Addenbrooke's Cognitive Examination for dementia [51], a paper based and easy to administer screen, currently in its third iteration (ACE-III) [52], though it takes substantially longer to complete than MoCA and MMSE. The ACE-III covers five cognitive domains; attention and orientation, memory, fluency, language, and visuospatial abilities [53].

The MoCA specifically, which was based on the MMSE, is a 30-point brief domain-general assessment of cognition, with sub tests covering visuospatial construction, executive function, language, memory, numeracy, working memory, and attention; it was designed to detect mild cognitive impairment and dementia [37]. The MoCA is the clinical standard for cognitive screening based upon historically high use in stroke settings [45, 47, 49]. For the visuospatial component of the MoCA, participants complete a small trail making test [54], alternating connecting numbers to letters in ascending order (e.g., 1-A-2-B etc), then draw a figure (e.g., a line drawing of a cube), and then draw a clock with the hands in and the numbers at a set time (e.g., 10 past 11). For the language component, participants name three pictures, repeat two sentences, and then are given a verbal fluency task to name as many words that begin with a specific letter (e.g., F). For the attention component, participants repeat two lists of numbers in the forward and backward order respectively, then tap for the letter A when the assessor read a long list of letters, and then they have

a serial subtraction task. For memory, the participants repeat a list of 5 words twice after they are read aloud by the assessor, then after a delay, the participant freely recalls the words. If the participant does not freely recall all words, they are given category cues by the assessor for missing words, and then are given multiple choice options for further forgotten words. Finally, the MoCA asks 6 questions for orientation (e.g., what is current date etc). The outcome of the MoCA is a single score out of thirty which ignores non-freely recalled words, and it has multiple acceptable cut offs with no subtest normative data. A cut off of <23 can be used closer to the acute event and beyond [55], but differing cut offs, including a cut off of <26 as originally prescribed [37], are widely used [56, 57].

Furthermore, the MoCA has mixed sensitivity and specificity evidence, often resulting from low sample numbers that produce the estimates or from demographically homogeneous groups (i.e., small age ranges, highly educated etc) [58]. Moreover, the MoCA was designed as a screen for mild cognitive impairment in older adults, not for stroke specifically which may affect subsequent sensitivity/specificity analyses in stroke [59]. Some 40 or so research articles have attempted, across multiple cultures, to generate demographically adjusted MoCA cut off scores, but the evidence base is not sufficient to support these cut offs [58]. A recent large scale study used 800+ healthy Dutch participants covering a wide age range to determine regression based demographically adjusted cut offs for the Dutch MoCA [58]. Kessels et al. [58] created tables of percentile scores for the Dutch MoCA across age group category (18-80), gender (man or woman), and education (low, average, and high); the first project of this kind at a large scale. Additionally, Kessels et al. [58] encouraged the use of consensus based clinical labeling of participant scores from Guilmette et al. [60], to inform judgement of MoCA scores. That is, whether a participant has scored *exceptionally high* or *average* or *exceptionally low* etc dependent on scoring category [58]. In this way, although the MoCA is the clinical standard, it has not got the widest validity evidence for sensitivity and specificity until now, and only in Dutch speakers.

Screening tests must have their validity examined in terms of criterion validity, at least via convergent validity and predictive validity among other forms [21, 61]. Criterion validity is an umbrella term for capturing elements of a construct under measurement (covered by convergent and predictive validity terms). Convergent validity refers to whether a construct is still captured, when measured in different ways; and it is measured by associations between different tests via correlations, multi-trait multi-method matrix, and latent modelling [62]. Predictive validity refers to whether a score on a given test predicts future outcomes which are relevant to the construct being measures; and is measured often with forms of regression modelling [62].

The MoCA and MMSE have had convergent, criterion, and predictive validity established in stroke samples [63] which is sufficient to establish them as validly measuring domain-general cognitive impairment. A 2012 study administered the ACE-revised (ACE-R), the MMSE, and a neuropsychological battery to 101 stroke survivors on hyper-acute and acute stroke wards and found limited sensitivity and specificity of both tools [64]. In a 2018 Cochrane review, post-stroke cognitive impairment measured with the ACE-III or mini-ACE was found to have insufficient sensitivity and specificity, with a small evidence base [65]. Thus the ACE is not preferred for stroke.

The MoCA and earlier MMSE, both suffer from an inability to provide a breakdown of specific cognitive domain impairments (e.g., deficits in language, memory, or attention etc) [38]. Whilst these subtests are clearly present within the tasks, there is no normative data for each subtest. I did reach out to various MoCA authors and the organisation administering licensing and all refused to provide subtest norms. While the subtask data must exist due to the volume of papers using and norming both the MoCA and MMSE, they both lack subtask and population specific normative data which would aid in detecting subtle changes in cognitive domains in different populations and across time [66, 67]. Both tests instead provide a single cut off for impairment, which is against clinical guidelines recommendations [41] which requires a demonstration of cognitive impairment in a

domain-specific way. The ACE-III falls more in line with clinical guidance than the MoCA or MMSE due to providing both an overall normed score and domain-specific subscores. All three cognitive screens do not, however, cover the stroke specific domain of praxis, nor visuospatial neglect explicitly. Similar in part to the MoCA and MMSE, the ACE-III does not require specialist equipment to administer, and it does not require specialist training, it is also free to use. As of 2021, the MoCA does require paid for specialist training to access materials and administer the MoCA legally, the MMSE is also costly and not freely available. This may not be the case for all licenses, for instance the local Oxford University Hospitals NHS trust does not have to pay, however, Oxford Health NHS trust does, and so we are prohibited from administering the MoCA on site. According to the latest clinical guidance, no screening tests should be used which do not provide freely available training for clinicians [43]. As the MoCA requires paid for licensing to administer and interpret, usage of this test is against current UK guidelines.

Moreover, all three tools were not developed for use with stroke survivors who commonly have language and perceptual/attention impairments [39, 68] which confound task performance on screening tools developed primarily for Alzheimer's disease dementia, focusing in particular on verbal memory [39, 66, 69, 70]. It is possible that due to the reliance on language in the tools, that stroke survivors with spared language abilities may mask their cognitive impairments in other domains, thus remaining undetected on screening and not getting targeted help [70]. While the tests are confounded by these highly prevalent stroke-specific impairments, the MoCA is still one of the most commonly used cognitive assessments tools in stroke care [45, 47]. Following from the Standards for Education and Psychological tests, the MoCA, MMSE, and ACE are not to be used on the basis of not being designed for stroke survivors, and not accounting for their different requirements, and therefore cannot be validly used. Although, the MoCA and MMSE certainly do save time with their brief administration times.

On the other hand, since inception new studies have supported some use of the MoCA in stroke. Note, however, that studies can be highly selective. For instance,

one recent study on 65 Taiwanese stroke survivors tested the validity of the MoCA against the Stroke Impact Scale [71], but excluded any stroke survivor with cognitive impairments on the MMSE [72]. Thus, the validity of the MoCA was tested in those likely to be spared in many cognitive domains. Moreover, in the same study the researchers found an association with the language/communication subscale of the Stroke Impact Scale [72]. Note that the MoCA total score and the subdomains of naming, attention, language, and abstraction also correlated moderately to highly with the language subscale. I interpret this finding to provide evidence against the use of the MoCA as it clearly related to langauge abilities even in those without severe cognitive impairment. A final note for the Wu et al. [72] study is that the MoCA total score and subscales did not significantly associate with many other subscales of the Stroke Impact Scale. Other highly cited research has claimed validity in stroke but fallen foul of the sample sample issue. For instance, one study entitled *'Montreal Cognitive Assessment and Mini–Mental State Examination are both valid cognitive tools in stroke'* by Cumming et al. [63] found good agreement between the MoCA, MMSE, and a neuropsychological battery in a small, but statistically powered, stroke survivor cohort; yet selected only those with preserved language abilities who were not so affected by their stroke that they were unconscious on admission. There are further issues with this seminal paper, however, one is that whilst they excluded those with language impairments, they otherwise were inclusive in their recruitment at an acute stroke ward in America, yet in 2.5yrs only managed to find 104 stroke survivors eligible, which does not seem plausible from experience.

The Oxford Cognitive Screen (OCS) [38] on the other hand, is a paper and pen based brief domain-specific cognitive screen designed to screen for impairments in five domains; language, attention and executive function, memory, numeracy, and praxis [38] and aligns with clinical recommendations and guidelines as a first line screening tool post-stroke [46, 73, 74]. These domains are referred to as *stroke-specific* domains in that these are the most affected cognitive domains post-stroke. In contrast to the previously mentioned screens, the OCS was designed from inception as a stroke-specific screening tool. Thus, it was developed with stroke survivors

who may have language impairments in mind. The main immediate benefit of the OCS, therefore, is that it directly follows clinical guidance, is domain specific, and stroke-specific. The OCS includes 12 subtasks (i.e., picture naming, semantic understanding, orientation, visual field assessment, sentence reading, number writing and calculations, cancellation, praxis, delayed free and cued recall of the sentence reading task, episodic recognition, and executive function). Each subtask can be classified as impaired or spared, aiding clinician interpretation of performance. OCS starts with in step tasks which inform how the rest of the OCS is administered. For the first task, if participants are unable to freely give the name of the picture aloud, they are offered written multiple-choice options for further tasks in the OCS to avoid communication difficulties impacting performance in other cognitive domains. Another adaptation is that if unilateral hemispatial neglect is detected and/or a visual field deficit is found, the remainder of the OCS is presented in central space at the stroke survivors midline, facilitated by the assessor, to avoid the confounds of either or both impairments [38].

The OCS has been found to be more sensitive to detecting cognitive impairment in stroke than either the MoCA or MMSE [39, 68] and very recent evidence supports use of the OCS over the ACE-III in stroke [75]. Note however, that it is possible greater perceived sensitivity comes at a cost of potential false positive findings affecting results in comparisons. However, as the OCS covers the five domains in a way that attempts to account for stroke-specific confounds such as language, neglect, and praxis, and is more sensitive and specific than both the MoCA and MMSE, the OCS is superior to three most commonly used tools. Although the ACE-III covers five domains as well, the superiority of the OCS is evidenced by its coverage of relevant stroke domains such as number processing and visual neglect; and its stroke-specific nature meaning it was designed to reduce confounds of language and perceptual impairments, where the ACE III heavily relies on intact language abilities. Furthermore, the OCS requires no specialist training to administer, and it is completely freely available for publicly funded research and clinical use via license, rendering the tool highly useful for stroke clinicians who are not trained

in cognitive assessments. Note, features such as keeping the materials at the midline are clearly noted in the OCS manual and test training/education. Using the validation framework presented earlier, the OCS has been designed with stroke specific difficulties in mind, thus is fit for use with stroke and have been validated as measuring the intended constructs [38].

In summary, cognitive screening is an important step in a stroke survivors journey which higlights both strengths and weaknesses they have following their stroke. Cognitive screening is required by best practice clinical guidance, yet there are several challenges clinicians face. Clinicians, who on the whole in stroke care, are not specialist and do not have neuropsychological support, are expected to select the best clinical tool available per situation, become trained in that tool, and interpret the tool. Currently, the most appropriate tools have not been historically used in clinical practice meaning many stroke survivors have been inappropriately screened.

0.2.1 Impact of the COVID-19 pandemic on gross-level cognitive impairment

The current sub-section remains on the topic of cognitive screening of primary impairments, but diverts the discussion into the impact of the COVID-19 pandemic. The pandemic fell at the beginning of my DPhil, and as such, impacted the first year of my studies in many ways. One impact was regarding translating the OCS for remote usage. This section details the impact of the pandemic specifically on already mentioned cognitive screening tools, and how I aided in the translation of the in-person OCS.

The OCS is commonly used in standard practice in stroke units across the UK and increasingly across the world, with over 2,000 licenses released for use in clinics (see www.ocs-test.org) and many international translations [66, 76–84]. However, OCS was unusable due not being validated for remote administration prior to or during the corona virus disease 2019 (COVID-19) pandemic. The COVID-19 pandemic required a shift in the modality of administration of screening tests, from being face-to-face, to remote assessment due to the inability to be in close contact

with others [85, 86]. As such, multiple screening tools were administered over phone or video teleconferencing without appropriate standardisation or validity [87]. Some cognitive screens have been validated for remote assessment, such as the telephone Montreal Cognitive Assessment (Tele-MoCA/T-MoCA) both pre-pandemic [88] and during the pandemic [89, 90], the Telephone Interview for Cognitive Status [91], and the 26-point telephone version of the Mini Mental State Examination (MMSE) [92]. Notably, these remote adaptations do not include stroke specific tasks that cover stroke specific deficits. Although the need for validated remote cognitive tests has grown, remote versions of the MoCA and MMSE are not easily accessible, requiring a paid license and paid training to gain access to materials. As previously discussed, clinicians must decide the best tool to use, without proper guidance, and a paid for license may discourage usage of the validated tools and could lead to using invalidated free versions. The OCS, however, has some evidence for its validity in a remote form.

As part of this thesis an empirical pilot investigation is presented into whether administering the OCS over the phone and/or via videoconferencing is equivalent to in-person administration. This pilot investigation was completed during the pandemic when in-person testing was heavily restricted. Please refer to chapter 1.

0.3 Assessment of complex and/or subtle cognitive impairments post-stroke

The focus so far of the introduction has been on the classification of stroke via gross level cognitive impairments and how they are screened in acute settings and remotely. Many stroke survivors, including those with mild strokes or Transient Ischaemic Attacks (TIAs), however, may present as unimpaired on structured cognitive tests yet have disabling impairments which prevent return to activities of daily living post-discharge [50, 93, 94]. Because these individuals are spared on gross level cognitive screens, and as cognitive screening is often the only cognitive testing received, these individuals may not receive an appropriate care package for their needs [95]. This

is because without an objective impairment classification, it is difficult for health care professionals to justify costs of care or to refer to an appropriate rehabilitation service [94]. When following up 45 stroke survivors with TIAs or 'non-disabling' strokes at 1-month and 6-months post stroke, it was found that approximately half had impairments in their activities of daily living as well as symptoms of ill mental health and lower quality of life [95]. In addition, half of the participants who were discharged and classed as fully recovered acutely, required rehabilitation post-discharge [95]. It is important to ensure quality of life and fullest participation in daily activities for stroke survivors, and to ensure that we are capturing even subtle impairments in stroke survivors acutely. As per new clinical guidance, further specific testing for those who could return to instrumental activities of daily living for jobs (e.g., finance, director roles etc) is recommended [43].

There are existing cognitive screens for subtle deficits following stroke. The previously mentioned MoCA, has been discussed in the context of primary/gross level impairments, but not yet in subtle impairment detection thoroughly. The MoCA was intended to be able to detect mild cognitive impairments [37]. Indeed, Pendlebury et al [96] found that the MoCA was able to detect mild impairments in those who are otherwise spared on other cognitive screening tests such as the MMSE. As often the case, high sensitivity may come at a cost of lower specificity. In the original investigations comparing the sensitivity of MoCA and MMSE in participants with mild cognitive impairment and separately the specificity in healthy older adults, the sensitivity and specificity was 90% and 87% respectively for the MoCA and 18% and 100% for the MMSE [37]. The lower specificity is commonly reported in clinical samples. The MoCA has lower sensitivity in stroke than the OCS (33.33% specificity, as calculated from 310 participants in Table 2 of Demeyere et al. [39]), and similar ratings were shown in cardiovascular participants, where the specificity for the cut off of <26 was 29-30% [97]. These participant groups are expected to have greater impairment, yet the MoCA was likely to miss the majority of those who were not impaired. The low specificity is less the case for milder strokes or TIAs. One study compared the MoCA and MMSE in 239 participants with mild stroke and TIAs and

followed them up 3-6 months later and completed a neuropsychological test battery; they found that the MoCA had classically high sensitivity to detecting impairment (88%) and reasonably high specificity too (67%) using the <26 cut off [98].

Overall, it is likely that the MoCA is a reasonable test in mild strokes/TIAs -without cardinal language or motor impairments- and healthy ageing to detect mild cognitive impairments, but it fails to detect impairments in stroke specific domains (for example spatial neglect or praxis) which may be gross domain-specific impairments. An ideal test, therefore, would be one that can detect domain-specific and domain-general subtle and primary/gross level deficits in a way that also captures stroke specific deficits. The Oxford Cognitive Screen - Plus is a an extension of the OCS which was designed to better capture subtle cognitive impairments to the same or better level as the MoCA, though keeping in mind the specific stroke population it is designed for, and would replace the MoCA as a screen for milder and gross cognitive impairments in stroke populations [50].

The OCS-Plus is made up of 10 subtasks with multiple normed outcome metrics from specific subtasks. The OCS-Plus includes the following tasks: picture naming, semantic processing, orientation, word encoding, a trail making test, a word recall task, a figure copy task, a rule finding task, and a cancellation test. The OCS-Plus has initial psychometric validity evidence from 320 healthy, highly-educated, older adults across different sites in the UK and in two sites in Germany [50]. The participants were administered the OCS-Plus, and in Germany, a subset of 86 were administered a neuropsychological battery including independent, standardised, and validated assessments that were matched by construct to each OCS-Plus subtask. The neuropsychological battery included the mentioned Addenbrooke's Cognitive Examination Revised (ACE-R) [64], the Consortium to Establish a Registry for Alzheimer's Disease Plus (CERAD) [99], Rey–Osterrieth Complex Figure Test [100], the star cancellation test taken from the Behavioural Inattention Test [101], and the Beck Depression Inventory [102]. Both the ACE-R and the CERAD Plus include all of - or elements of - the MMSE. The project found that the OCS-Plus achieved benchmark correlations for convergent validity and showed, per subtask,

near perfect divergent validity [50]. Following the Standards for validity evidence discussed previously [21], the OCS-Plus was designed for stroke survivors and has construct validity evidenced in a healthy adult sample.

There were, however, a number of limitations to this original validation. The OCS-Plus was not initially compared to another standard full screen (e.g., the MoCA) that is commonly used in stroke practice, for the detection of subtle cognitive impairments other than through correlations per subtask, even though the MMSE makes up part of both the ACE-R and the CERAD Plus batteries. This meant that although the OCS-Plus was intended to detect subtle impairments we do not know if this happens at a similar, worse, or better rate of detection as other competitor tests like the MoCA. More specifically, the comparison tasks per subtask were not well matched. For example, the trail making test in the CERAD Plus [99] is based on the original trail making test [54] which included alternating between numbers and letters, and as such it is language and number laden. In the OCS-Plus the trails is a shape analogue, requiring alternating connection of large squares and small circles. While the CERAD Plus trails does measure the same construct as the OCS-Plus trails, hence correlations found between them, the added cognitive load of number and letter processing may attenuate the evidenced construct validity of the OCS-Plus trails. Instead, a version of a trail making test which also removes language and number processing requirements would be a better comparison to the OCS-Plus trails, and if correlations are higher it would likely reflect the removal of language/number confounds.

Moreover, the original validation omitted to validate the rule finding task from the OCS-Plus, and it was not discussed in the paper, nor clear why this was done. As such, there remains a test in the OCS-Plus with unknown validity evidence. Moreover, the validation was only completed in healthy and relatively highly educated controls, while this is useful for generating normative data, it inevitably leads to some ceiling effects on easier tasks. This was the case with the picture naming task, where healthy controls achieved 100% accuracy, and as such the range in scores reduced the reliability estimates for the test and single errors threw

correlations due to acting like outliers [50]. There is a need therefore to re-validate the OCS-Plus test in a sample which would not have ceiling effects, would have different education levels, and to compare the OCS-Plus to a test like the MoCA to establish sensitivity to detect subtle cognitive impairments. Again, following the Standards, both construct and respondent validity evidence is needed here.

In summary, this section discussed the assessment of of subtle or mild cognitive impairments, again discussing the strengths of the MoCA and OCS and now OCS-Plus in relation to each other. The MoCA is consistently reported on throughout this thesis as a comparison, hence its prevalence in the introduction. A specific comparison is made between the OCS-Plus, a new neuropsychological battery, and the MoCA, in chapter 2.

0.4 Ecological validity of assessments of cognitive impairments post-stroke

The final section of both the introduction and my thesis, is the relationship between cognitive screening and real life behaviours and performance in everyday activities.

The role of the neuropsychologist or assessing clinician has changed over the last century from diagnosing disorders arising from brain injury or disease to assessment of cognitive abilities relevant to everyday life, including ability to work and go to school [103]. However, the diagnostic tests have not adapted alongside this shift to reflect real life behaviours. Over 40 years ago, more research into neuropsychological tasks that reflect everyday behaviours was called for, using reliable estimates of everyday behaviours that were comparable across samples, and using a diverse range of participants that were not homogeneous in demographics [104]. A long-standing debate in psychometrics more generally is around ecological validity [105, 106], that is, whether there is a clear link between performance in a controlled environment (i.e., neuropsychological assessment or cognitive screen) and a naturalistic environment (i.e., real-life behaviour) [107]. The debate is specifically pertinent regarding higher order functioning such as executive functions. Following

2023 recommendations for standardisation of post-stroke executive function research [108], executive function is a higher level umbrella term encompassing many complex cognitive abilities such a planning, switching attention, inhibition of preponent response and updating of information [109] and historically has been related to the frontal lobes among other regions [109–111]. The frontal lobe paradox challenges the ecological validity of traditional neuropsychological tests of executive functon as it refers to the observation of patients who are classed as unimpaired on classical structured neuropsychological assessment, but when given open ended complex tasks related to real-life behaviours, they show impairment [106, 112]. The existence and concept of the frontal lobe paradox is debated currently, due to its basis being small scale case study research from prior to the year 2000 [113]. More appropriately, the frontal lobe paradox can be referred to as a strategy application disorder [114, 115]. Nonetheless, whether there is a direct link between frontal lesions (or other connected or disconnected regions of the brain) and impairment in either controlled tasks or the naturalistic environment, there are individuals who pass neuropsychological tasks but cannot successfully complete goal-driven activities in their daily lives.

Traditional neuropsychological tests such as the trail making test (TMT) [54] -which is a popular test of executive function- [116] for example, were designed to capture cognitive flexibility and diagnose issues in this construct [54] and it was not intended as a test used to capture everyday behaviours. The trail making test shows *veridicality* which is the degree to which tasks are statistically related to real life functioning [117], as opposed to *verisimilitude* which is where the demands of a neuropsychological test reflect the demands of real life [103], and the trail making test does relate to some everyday behaviours [104]. For instance, performance on the trail making test associates with criminal activities which are related to shifting tasks [118, 119]; showing that the trails task has at least veridicality. However, it is not directly clear what aspects of life neuropsychological tests reflect or whether they can reflect all aspects of everyday behaviours at all. Moreover, there is the question as to whether it is important, clinically or neuropsychologically, to try to reflect all aspects of life at all.

Current established and standardised neuropsychological tests lack face validity. All the screening tests like the MoCA, MMSE, ACE-R III, and OCS, were all designed to measure cognitive constructs, not to directly map onto real life behaviours. It is not the intention of any of these screens to reflect everyday problems, however, with the long-standing and increasing desire to detect problems in everyday lives as mentioned at the start of this section, it is still relevant. Some evidence has linked these screens, which lack verisimilitude, to everyday behaviours. For instance, the MoCA has been shown to have a weak to moderate prediction of activities of daily living over and above other demographic variables [120], as well as the MMSE predicting higher odds for functional dependence [121]. Further, the OCS taken acutely can predict activities of daily living in in-patient stroke survivors, particularly impairment in executive function [122], and can predict functional recovery of stroke survivors over demographics and stroke severity [123]. There are other tasks, however, which have verisimilitude but are still structured enough to have standardised administration across participants and rigorously test cognition in a way that shows veridicality.

0.4.1 The Multiple Errands Test

One such test is the Multiple Errands Test (MET), which was created to assess realistic goal-orientated behaviour in a standardised and quantifiable way [106]. It aims to mitigate the disparity between neuropsychological assessment and everyday life. The mitigation, in the simplest terms, is achieved by good task performance requiring good prospective memory, subtle planning abilities, and ability to cope in an ill-structured scenario [115, 124]. This subsection will detail the MET, its adaptions, and new versions for which validity evidence is presented in this thesis.

The MET, now adapted into multiple versions [124–128], asks participants to shop for set items, normally within time and cost limits, and to answer questions. The task is most often carried out in a real shopping centre [106] or hospital [125], and thus examines behaviours in the most realistic way with similar demands of multitasking as in everyday life. The task assesses many aspects of executive

functions and wider cognition, including planning, switching attention between tasks, inhibition in the form of perseveration of behaviours or inability to stop proponent responses and self-monitoring [129, 130]. There are five characterisations of mutli-tasking (accordidng to Burgess, [114]); 1) many tasks have to be completed, 2) tasks are overlapping to increase efficiency, 3) one task can be completed at a time, 4), distractions and interruptions will appear, and 5) the requirement for multiple tasks to be running and the end of a given task is not signaled by the task itself. These features appear in the MET through the length of tasks, delayed timed tasks, the number of tasks and their overlap, and the setting in which the MET is run leads to interruptions or distractions.

In the original MET and some adaptions, participants are faced with a large load on memory, by having to memorise a shopping list and rules to follow [106]. The MET is widely praised for the ability to standardly assess behaviour in an ecologically valid way and is a preferred test by occupational therapists [131]. One newer version of the MET is the MET-Home [131]. Early MET adaptations were suitable for hospital settings and virtual settings (see examples Alderman et al. [125], and Jovanoski et al. [127]), but the MET-Home was the first to provide a standardised approach to home testing. Note, the much shorter 'Home-MET' was also introduced in 2020 with a Chinese mild cognitive impairment and healthy ageing sample [132].

The MET-home [131] comprises of 14 tasks to be completed by a participant in their own home. The MET-Home is relevant for chapter 3. Tasks in the MET-Home were chosen to typically require movement between rooms where possible and to use different modalities. For example, one task requires finding a number and calling someone, another task requires looking something up, and another task requires getting an object. Participants are scored by an in-person examiner who assess whether each task was completed accurately, partially accurate, or omitted entirely. Further, the examiner records the number of passes by places where a task could have been completed, and any number of inefficiences the participant displayed (e.g., taking a call part way through the task and getting distracted, or feeding a pet etc). This is a particularly relevant MET as it can be run in a stroke survivors home

without the need to travel to hospitals, shopping centers, or locations with virtual reality set ups. Scoring of the MET-Home, as relevant for this thesis, is presented in the appendix. The MET-Home and the Home-MET differ in that the Home-MET is shorter, and has more requirement for visiting different locations, with specific mention to go to certain locations. This requirement to go to specific locations lessens the use of the Home-MET as it restricts usage to those with homes suitable for the assessment. The MET-Home is designed to be used across multiple rooms, but if an individual lives in a one room studio, the MET-Home is still suitable, whereas the Home-MET is not. A specific example from the Home-MET is a task requiring participants to affix a calendar to the door of the kitchen, which is highly relative to the individuals home layout and ownership of a calendar.

However, the administration of the MET, including the MET-Home nor Home-MET, is not appropriate in acute settings, nor for many stroke survivors who are unable to walk or shop for themselves, but who may have to make complex decisions requiring executive functions in their lives [130]. Versions were made in the early 2000s which sought to remove the burden of patient transport and assessment, but still they required high mobility levels to complete, which many stroke survivors do not have (see example Knight et al. [124]). Indeed the sample used to validate the simplified MET hospital version [124], were intelligent chronic head-injury survivors with no language or perceptual difficulties, thus not a representative sample to base a simplified version on. In a qualitative study examining the views of occupational therapists on the MET and MET-Home, it was found that while the MET was praised, concerns were raised around safety and time/money costs [131]. Moreover, in a clinical context, a MET is often not always feasible. Most prominently, the need to take patients out into the real world raises practical concerns regarding transport, staff time, and patient safety. Several versions of the MET have since been created and a 2020 systematic review summarised the psychometrics of the MET [129]. Real-world versions have been adapted for patients in hospitals [124, 133], shopping centers [125] and large stores [134], as well as a home-based version [126] addressing some of these barriers. However, patients with neurological conditions

often have co-occurring motor impairments restricting their ability to complete the task. Virtual reality/computerised versions may be able to provide a solution here, and several have been developed (e.g. Cipresso et al. [135]; Jovanovski et al. [127]; Rand et al. [136]; Raspelli et al. [137]), though these often bring high costs and needs for technical equipment and expertise which are often not readily available, or at least were not available for 30 years since the test was designed.

In a 2020 systematic review of MET literature [129], 33 articles that reported a version of the MET were found and their psychometric properties were assessed. The MET was commonly scored by accuracy of task completion, task omissions and partial omissions, as well as scored regarding rule breaks [129], with partial omissions being most sensitive to executive impairment [133]. The review found that for many versions of the MET, there was generally good internal consistency, good inter-rater reliability, sufficient test-retest reliability, good to adequate convergent validity, and good ability to differentiate clinical groups.

With regards to convergent validity, the MET has been found to converge with a variety of standardised neuropsychological measures of executive functioning. For example, the performance on the MET was found to correlate with the Trail Making Test Trail A time and Trail B, and Trail B/A performance [125, 127], and with inhibition impairment [105]. A 2014 systematic review summarized nine papers that included the MET, validated measures of executive function, and participants with acquired brain injury [138]. The review showed that each paper used a diverse range of neuropsychological tests, commonly including subsets of the Behavioural Assessment of Dysexecutive Syndrome (BADS; Wilson et al. [139]), digit span tests, fluency tests, figure drawing tests, story recall tasks, attention tests, and activities of daily living assessments. Convergent associations between the performance on the MET and the following were found: the Modified Six Elements Test [127] and Zoo Map test [140] of the BADS (although note no evidence of convergence has been found in other research [141, 142]), Rivermead Behavioural Memory Test [143], Comprehensive Assessment of Prospective Memory [144] and Instrumental Activities of Daily Living Scale [145]. Evidence for divergent validity would result

from comparing the MET to non-executive measures [129], such as memory and intelligence [146]. Explicit testing for divergent validity has however not been common practice in studies on the Multiple Errands Task. The systematic review by Rotenberg et al. [129] highlighted that no examinations of divergent validity were conducted in the studies included.

Previous versions of the MET have failed to consider placing the MET within the framework of a feasible, short, and ecologically valid screening tool that is inclusive of the clinical reality after acquired brain injury. Up to now, virtual versions of the MET have used joysticks [147], virtual reality wands [148], and desktop keyboard set ups [149], but are often too long to complete in clinical settings and require complex set ups. These can be expensive, even if the cost of virtual reality technology has decreased in recent years [150]. Outside of the Multiple Errands Tests, other short versions of ecologically valid tests of executive function have been developed, including for instance the 15 minute Hotel task [151], which is similar to a six elements task [106]. This example is a table top task, which requires a set-up of complex materials and a skilled examiner for administration and scoring. The computer-tablet is an alternative format not yet tried with the MET that is able to address these issues. A computer tablet app version of the MET could provide guided administration, remove the complexity of any material set up, automatic scoring and shorten the time necessary to test.

0.4.2 The Oxford Digital Multiple Errands Test

A simplified computer-tablet version of the MET test was developed, named the Oxford Digital Multiple Errands Test (OxMET), to be run at bedside as a screening tool [130]. This subsection details the OxMET specifically, which I co-designed as part of my fellowship and I present the manuscripts relating to this test in chapter 3.

The OxMET is a simplified computer tablet MET designed to be used by those with cognitive and motor impairments [130]. It achieved the aims of a MET by having multiple tasks to complete within time constraints but there are significant departures from the original MET and subsequent adaptations. In

alignment with other METs, OxMET requires participants to buy six items, to answer two questions, and to do while following strict rule and time constraints. In difference to other METs, the OxMET does not have multiple tasks which must be completed concurrently, indeed all tasks can be completed one-by-one in any order at all. Thus when referring back to the definition of multitasking presented earlier [114], the OxMET does not meet criteria two, to have overlapping tasks which can be done in such a way to increase efficiency of all tasks being completed. Otherwise, the OxMET still meets the criteria of being a multitasking test as it requires participants to complete multiple tasks, to do so one at a time and it presents distractions in the form of distractor shops and distractor items. It could be argued that the design to not have planned interruptions to distract the participant means OxMET does not fulfill criteria five. The OxMET was designed in such as way as to capture core deficits in executive functions that would affect everyday behaviour, in line with METs as a whole.

The OxMET has been psychometrically assessed in 124 healthy controls and 105 stroke survivors for internal consistency, test-retest consistency, convergent and divergent validity, known-group discriminability (healthy ageing and chronic stroke) and it has had its sensitivity to general impairment assessed in comparison to other tests [130]. The OxMET was compared to the OCS-Plus [50], as well as two standardised and validated tests; the Zoo Map subtest of the Behavioural Assessment of Dysexecutive Syndrome [139] and the Pill Box test [152]. The OxMET has promising psychometric qualities showing evidence for its use as an assessment tool for executive dysfunction.

In terms of other validity evidence, however, the relation between functional outcome measures (measures of everyday behaviour) and the MET in general is mixed and no evidence is present for OxMET. It is important to examine functional outcomes as this is what clinicians use in practice to reflect real life behaviour in everyday scenarios. In the systematic review Rotenberg et al. [129] found there 12 studies that examined the relation of the MET in general to everyday behaviour with 24 hypotheses examined from these papers. There were inconsistent relationships

between METs and everyday behaviour, with 11 of 24 hypotheses being confirmed [129], and mixed evidence being found even when using the same measure across studies. It is possible that the measure used to represent real life in these METs presented in the systematic review was not the most ecologically valid measure. The measure used most often was the dysexecutive questionnaire (DEX) [139]. Thus, while the face validity of the MET in general, and the OxMET, is good, its predictive validity of real life is not supported. Meaning the test has the opposite problems to the cognitive screens discussed in the previous section; the MET has evidence for verisimilitude, but not veridicality. What is needed is further research to establish how performance reflects real life everyday behaviour. Health professionals would then have an evidence base behind the MET, among other neuropsychological tests, so they can use performance to inform clinical decision making in practice.

In summary for the whole section on ecological validity, the switch to relating impairments on screens or assessments tools more broadly to everyday behaviour has lead to a necessary shift in how we design tests. It is now not only important to design tests with the intended participant in mind and that measures the right constructs, but to do so in a way that reveals something about the respondents behavior outside of the testing situation. In chapter 3 I present two separate studies examining different aspects of ecological validity of the OxMET, and an additional study which provides normative data for test that has both versimilitude and veridicality is presented separately in the appendix.

0.5 Thesis structure

Thus far, the standards for establishing validity evidence for cognitive assessments have been discussed, in context of how we define and classify impairments post-stroke. Assessments of primary and subtle cognitive impairments have also been described in contrast to each other with strengths and weaknesses highlighted between cognitive screening tools, such as the OCS and MoCA. Finally, the ecological validity of cognitive assessments, in context of modern practice to identify problems in daily

life rather than to diagnose a cognitive impairment, has been discussed. The key objective of this thesis is to examine new stroke-specific assessment tools and establish their validity for use with stroke survivors. The aim of the work presented is to provide health professionals with an evidence base to make judgement about which tools are best practice for use, by highlighting both strengths and weaknesses of each tool.

The thesis is divided into the following broad chapters which contain different published and in preparation manuscripts:

1. Remote cognitive screening
    - A published survey on how the COVID-19 pandemic affected assessment of cognition broadly
    - A published pilot validation study comparing in-person OCS to a remotely administered OCS in a chronic stroke sample
2. Screening for subtle cognitive deficits
    - A published comparison of the OCS-Plus in relation to a neuropsychological battery and MoCA in both subacute and chronic stroke survivors
3. Ecological and predictive validity of Multiple Errands Tests
    - A published study, which compared the OxMET and an in-person MET task standardised for the home
    - A published study examining the predictive validity of the OxMET to functional outcomes in a subacute stroke ward, and activities of daily living chronically post-stroke

The final chapter of my thesis will conclude with how each of the presented studies fit with the burden of validity evidence in new test generation, and how each of these address strengths and limitations of the current discussion on cognitive assessments. Final recommendations will be made for valid usage of each presented test in practice.

All analyses presented, and the writing of this thesis, was conducted in R Studio (R version 4.2.2 (2022-10-31)) [153] using the *oxforddown* thesis r package for formatting [154].

1

Remote cognitive screening

1.1 The COVID-19 pandemic altered the modality, but not the frequency, of formal cognitive assessment

This section is partially altered from the peer-reviewed publication: **Webb, S. S.**, *Kontou, E. & Demeyere, N. (2021) The COVID-19 pandemic altered the modality, but not the frequency, of formal cognitive assessment, Disability and Rehabilitation, doi: 10.1080/09638288.2021.1963855*

1.1.1 Abstract

Purpose I investigated the impacts of the COVID-19 pandemic on the modality of formal cognitive assessments (in-person versus remote assessments). **Materials and Methods** I created a web-based survey with 34 items and collected data from 114 respondents from a range of health care professions and settings. I established the proportion of cognitive assessments which were face-to-face or via video or telephone conferencing, both pre- and post-March 2020. Further, I asked respondents about the assessment tools used and perceived barriers, challenges and facilitators for the remote assessment of cognition. In addition, I asked questions specifically about the use of the Oxford Cognitive Screen. **Results** I found that the

frequency of assessing cognition was stable compared to pre-pandemic levels. Use of telephone and video conferencing cognitive assessments increased by 10% and 18% respectively. Remote assessment increased accessibility to participants and safety but made observing the subtleties of behaviour during test administration difficult. The respondents called for an increase in the availability of standardised, validated, and normed remote assessments. **Conclusions** I conclude that the pandemic has not been detrimental to the frequency of cognitive assessments. In addition, a shift in clinical practice to include remote cognitive assessments is clear and wider availability of validated and standardised remote assessments is necessary. Key words: COVID-19, cognitive assessment, remote assessment, video conferencing, telephone conferencing, cognition, teleneuropsychology

1.1.2 Introduction

In late 2019 a novel corona virus (COVID-19) was identified in China and spread to Europe within months [155]. After the first recorded case of COVID-19 appearing in the UK at the end of January, in March 2020, the United Kingdom implemented a national lockdown [156], limiting or restricting face-to-face contact for many clinical services. The rate of face-to-face appointments at general practices decreased exponentially, with 71% of routine consultations being conducted remotely, up from 25% the same time in 2019 [86].

In 2019, a mixed methods survey investigation of videoconferencing in clinical neuropsychologists in Australia revealed only 6.70% of 90 respondents reported using video conferencing for assessment, with the highest percentage of using videoconferencing reported for feedback and interventions, 18% and 20.80% respectively [157]. This is likely to be much higher than in other parts of the world, given that Australia already had relatively well-established telehealth services (e.g. Bradford et al. [158]). The same survey also found that where videoconferencing was considered, it was used very infrequently. The lack of engagement with remote or digitised assessment seems to exist despite reported high (98%) satisfaction with teleneuropsychological services [159].

During the pandemic, services needed to swiftly switch to remote and digital assessments with likely long-term changes taking effect [160]. Guidance was published on adjustments, for example, from stroke services [161] and memory clinics [162]. Guidance from the Royal College of Physicians suggested virtual assessments could be utilised such as WhatsApp or Facetime by multidisciplinary teams for virtual ward rounds and rehabilitation [163] in order to continue practice. Guidance from the Division of Neuropsychology encouraged remote assessment, but needed caution in interpreting results in comparison to existing normative data [164].

This swift change also encouraged the use of cognitive assessments that were not validated or did not provide normative data for remote assessment [160]. Encouragingly, when comparing remote and in-person assessments, research so far has shown performance to be similar across versions [165]. For instance the use of the Montreal Cognitive Assessment (MoCA) via video conferencing had not been fully standardised with normative data in 2020 [166], but it was recommended to use by the team behind the MoCA [167] given partial validation for its version [160]. Since publication of the current article in *Disability and Rehabilitation*, remote MoCA validations have been done [90]. I anticipated a rise in further validation research into the use of remote assessments, allowing clinicians to be more confident in their interpretations of remote assessment results.

In 2020, there were limited data regarding the frequency of use of remote assessments of cognition prior to and after the COVID-19 pandemic began in a global sample. One survey in late March 2020 surveyed 372 board-certified neuropsychologists in America, and found that, following the start of the pandemic, in March 2020 only 21.20% of neuropsychologists were conducting in-person assessments, and 66.70% planned to use, or already were using, teleneuropsychological services for assessments [168]. Of the 66.70% of respondents using teleneuropsychology, only 25.50% reported using remote assessments to administer neuropsychology batteries, with the majority reporting using the format for interviews (73.10%) and screening tests, verbal tests, brief batteries, dementia batteries, mental status exams (66.40%).

A separate survey study during the pandemic in the USA received responses from 230 independent practitioners of clinical neuropsychology and found that only 14% reported using teleneuropsychology for intake interviews, feedback, and some testing, and 3% for full neuropsychological examination [169]. I note that this survey only asked about full neuropsychological batteries, rather than whether individual neuropsychological tests were used, for example, only using an individual subtest.

In 2020, the teleneuropsychology surveys have exclusively recruited neuropsychologists or psychologists and none have explicitly included wider practice in cognitive assessments including cognitive screening. Here, I wished to acknowledge the breadth of professions that assess cognition in clinical practice, across a range of settings including neuro-rehabilitation and memory clinics. In this 2020 pre-registered study (registration available here https://osf.io/hvmx9) with full data and analysis repository (see https://osf.io/uv8jr/), I asked health care professionals from a range of settings about their experiences of assessing cognition remotely and face-to-face through a web-based questionnaire. I aimed to estimate and compare the rates of remote assessment pre-COVID-19 pandemic and post-COVID-19 pandemic, particularly focussing in on assessing the reported rates of cognitive assessments specifically (aim 1). I further aimed to identify which cognitive assessments were used both pre-COVID-19 pandemic and currently, with a subset of questions specifically focussed on the Oxford Cognitive Screen (aim 2). Finally, I set out to gather data on perceived barriers, challenges, and facilitators to remote assessment of cognition (aim 3). Similarly to the previously described online questionnaires on the use of teleneuropsychology before and after the COVID-19 pandemic, and as an extension of questions used previously [170], I anticipated lower use of teleneuropsychology prior to the pandemic, with an increase in phone and video conferencing since March 2020.

1.1.3 Methods

I report how I determined my sample size, all data exclusions (if any), all manipulations, and all measures in the study (21 word solution for transparency by Simmons

et al. [171]). I followed the CHERRIES Equator Network guidance [172] when reporting my web-based open-questionnaire. Approval for the study was gained from the Central University Research Ethics Committee (R51993/RE001). Informed consent was gained from respondents through a mandatory item on the web-based survey which should only be checked if the respondent was above 18 and agreed to take part, after reading the participant information sheet. In advertisements of the survey, respondents were told completion of the survey would take 5 to 15 minutes. The participant information sheet included all information regarding the ways in which data were stored, where, and for how long, who the investigating team were and the purpose of the study

Respondents

Two power analyses were conducted to establish minimum sample size to establish differences in remote cognitive assessment practice pre- and post-March 2020. The power analyses for both a one-sided paired t-test and a one-sided Wilcoxon signed-rank test (depending on data characteristics), were set with an alpha of .05, a small Cohen's d of .20, a paired sample, a one-sided test, and 80% power using the pwr [173] (version 1.3-0) package in R Studio [174]. The power analysis revealed I needed a minimum sample of 156 participants for the paired t-test.

I did not reach the expected sample size prior to the completion date set out in my pre-registration criteria, so I conducted a sensitivity analysis to see which effect size I could detect given the final sample of 114 respondents after exclusions, detailed later. With this sample size, I could detect with 80% power, an alpha of .05, using a paired and one-sided test, and a Cohen's d of .23.

I opened up the web-based survey for data collection on the 17th September 2020 and closed data collection on the 5th November 2020. The intended completion date was to coincide with the second national lockdown in the UK in November 2020, and I did not seek extension of the data collection period as I achieved an adequate sample size to detect a marginally different smallest effect size of interest. Respondents were convenience sampled by the research team via word

of mouth, social media, the lab websites, and email. In addition, I contacted the British Psychological Society Division of Neuropsychology, The Royal College of Occupational Therapists Specialist Section - Neurological Practice, various stroke wards, and occupational therapists to snowball the study. The inclusion criteria were that the respondents were aged 18 or over, willing to give informed consent for participation, and were healthcare professionals with direct contact with formally assessing cognition of service users and / or overseeing the assessment of service users. I note that where I have advertised the web-based survey has biased the sample to psychologists/neuropsychologists, and occupational therapists working in stroke services, over and above other professionals who conduct formal cognitive assessments in other settings.

Respondents who were not healthcare professionals or did not have any current experience completing or overseeing cognitive assessments with clients were excluded. The approved procedures did not allow for data collection through cookies or using the IP address of participants to track unique visitors to the web-based survey. There were 1451 link clicks to the survey landing page, of which 31 of those who followed the link continued through the consent process and then exited the survey, 20 of those who followed the link proceeded to survey page 3 and then exited the survey, 10 exited after page 4, and 3 exited after page 5. 136 visitors completed the survey and all mandatory items, and submitted their responses. Of those non-unique visitors who passed the informed consent page (1421), I had a completion rate of 10.44% (1421/136).

After exclusions, detailed in the appendix, I analysed data from 114 respondents. Further, I present the advertising materials, as approved through ethical review, in the Supplementary Materials

Materials

A web-based questionnaire was created and hosted on JISC Online Surveys [175]. The survey was designed in consultation with a clinical psychologist working in neurorehabilitation services and to ensure appropriateness and face validity of

the web-based questionnaire items for the intended purposes. In addition, the web-based questionnaire incidentally included items from a recent published web-based questionnaire [170] which was published shortly before the online pre-registration of the web-based questionnaire. The web-based questionnaire had a total of 34 items, of which nine were optional extras regarding the Oxford Cognitive Screen (OCS)[38]. All items of the web-based questionnaire are available from the data repository (https://osf.io/fxgqt).

Of the original 34 items (including sub-items), 19 items were available to all respondents. There were optional additional questions regarding the OCS and its use before and after March 2020, as well as whether it had been administered remotely as per recent published guidance by the authors (see https://innovation.ox.ac.uk/news/ocs-remote-administration/ and see https://tinyurl.com/yxz5hafn for demonstration). The final items about the OCS remote administration were to gather specific feedback on this tool.

I structured the web-based questionnaire such that I could fulfill the aims of the project, with the first few items concerning the rate of cognitive / other assessments pre-COVID-19 pandemic (i.e., pre-March 2020). The additional items, only available where the participant responded 'Yes' to whether they had conducted or overseen assessments of cognition (item 8) aimed to gather data for the use of remote assessments of cognitive functioning both before and during the COVID-19 pandemic. The final section enquired about perceived benefits, challenges, and facilitators of remote assessment of cognition, which was included to fulfil aim three. The optional section on the OCS was included to gather feedback for the authors on the remote assessment version of the OCS, which was released online in May 2020.

Procedure

The web-based questionnaire did not require an account to access and included the information sheet before beginning the web-based questionnaire, and consent process to take part. Respondents were allowed to exit the web-based questionnaire at any point simply by closing the web-based questionnaire tab or browser.

Respondents followed a link to the survey, and after reading the information sheet and consenting to take part, were taken to four screens with items in fixed order. All data were anonymous and not linked to names or personal information. Once completed, respondents could request a completion certificate to include in their portfolio evidencing continued professional development, or to confirm their interest to participate in further research. Respondents were asked to provide their completion ID to obtain their certificate. No linking information was stored following the sending of the certificate.

Data analysis

I explored descriptive participant statistics from the web-based questionnaire, including where respondents were based and what roles they had. Note, due to the majority of respondents being from the UK, I split the data into UK versus non-UK respondents, and present the main results from the UK sample only, and replicate results in the non-UK sample at the end of each section. I present results per aim of the study.

I conducted inferential analysis in the form of median comparisons of percentages of the modalities of assessment prior to and after March 2020 to examine whether there was a shift in format of assessment. The data for percentages of face-to-face versus teleneuropsychology violated the assumption of normality so I used Wilcoxon Signed Rank tests for comparison.

I counted the frequency of cognitive assessments listed by respondents before and after March 2020, and the percentage of responses which had reported changing assessments used after March. As the OCS was identified to be commonly used, and it was of interest to investigate, I summarise the findings from the optional OCS questions here. Finally, I summarised the perceived benefits, challenges, and facilitators.

In order to summarise the qualitative elements of the web-based questionnaire, a standardised coding system was developed, to establish the frequency of themes emerging from the data set that were shared between respondents. The standardised

Table 1.1: Reported frequencies of countries worked in.

Country	Count
United Kingdom	82
Canada	7
Belgium	5
United States	3
Australia	2
Ireland	2
Italy	2
New Zealand	2
Denmark	1
Finland	1
France	1
Norway	1
Pakistan	1
Portugal	1
South Korea	1
Spain	1
Sweden	1

coding scheme, as well as the associated frequencies, is available in Table 5.1 of the appendix.

1.1.4 Results

Sample Characteritics

The 114 respondents who passed the attention check were mainly based in the UK (82 respondents, see Table 1.1 for respondents per country) and worked in a variety of settings and held a varied number of roles. Figure 1.1 summarises proportions of respondents for each setting and role worked in the UK sample.

The 'Other' option for setting worked, selected 27 times, revealed multiple community-based settings (n=6), paediatric settings (n=3), neurological inpatient facilities (n=3), private practices (n=4), vocational workplace, a sleep lab, a radiology department, a physiotherapist, an outpatient psychiatric facility, a cognition clinic, an emergency department, an eye outpatient's facility, a child disability team, a general geriatric clinic, and a movement disorder clinic.

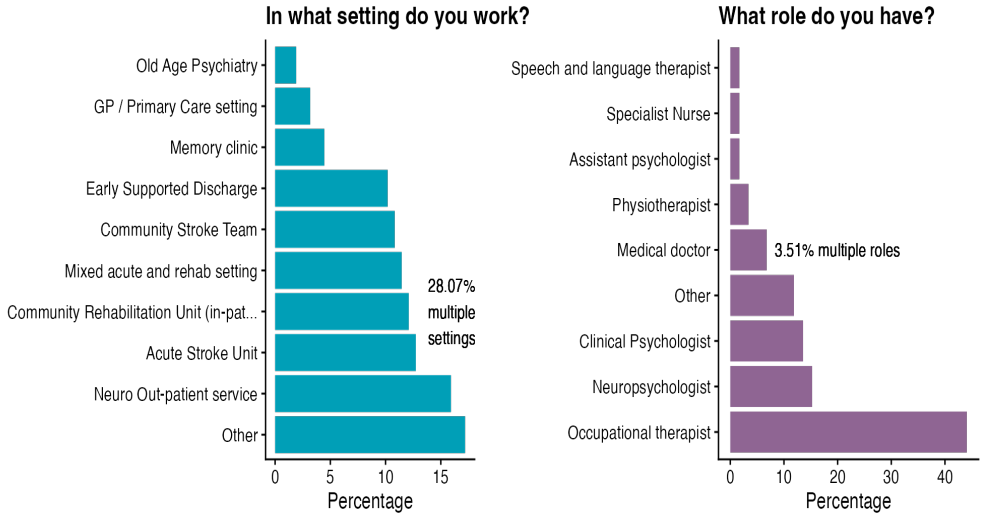


Figure 1.1: Percentage of participants who responded for setting worked, roles worked, and countries worked in items in the survey.

The 'Other' option for role, selected 17 times, revealed roles such as behaviour analyst, clinical researcher (n=2), therapists (n=2), a trainee clinical psychologist, a counsellor, a dietician, a nurse, an occupational psychologist, an orthoptist, a physio assistant, a radiographer, a receptionist, a sleep technician, a social prescriber, and a team leader.

The Rate of Remote Assessment (Any) Pre-COVID-19 Pandemic

In the UK only sample of 82 participants, 13.41% of participants used telephone calls and 1.22% of participants used videoconferencing assessment methods prior to March 2020, non-cognition specific. In the non-UK sample of 32 participants, 0% of participants used telephone calls and 9.38% of participant used videoconferencing assessment methods prior to March 2020, non-cognition specific.

Cognitive Assessments Pre- and during COVID-19 Pandemic period

Of the 82 UK respondents, five did not complete any cognitive assessments, and 10 respondents from the UK sample incorrectly completed items 9a-c (e.g., percentage/absolute values did not sum to either 100% or the number of assessments run) and had to be excluded in analyses on percentages of cognitive assessments pre-March 2020. Respondents reported conducting or overseeing a median of 10

cognitive assessments per month prior to March 2020 (IQR= 15, range = 1 - 400). Of these cognitive assessments, they reported 100% (IQR= 0, range = 100 - 100) were face to face, and 0% were via telephone or by videoconferencing.

During the COVID-19 pandemic, the median monthly number of cognitive assessments conducted or overseen by the respondent in the last two months was 6 (median; IQR= 10, range = 1 - 400). Of their conducted or overseen cognitive assessments, 72.53% (IQR= 50, range = 0 - 100) were face to face, 9.74% (IQR= 20, range = 0 - 100) were via telephone, and 17.74% (IQR= 20, range = 0 - 100) were by videoconferencing.

I conducted three one sided Wilcoxon signed rank tests with continuity correction, due to normality violations, with a Bonferroni alpha correction level of (.05/3 = .016), to compare the numbers of cognitive assessments administered before and after March 2020, as well as to compare the percentages reported for face-to-face, videoconferencing, and telephone cognitive assessments before and after March 2020 (see 1.2). Importantly, I found that there was no significant difference between the numbers of cognitive assessments completed before March in the average month, and in the last two months on average (MD= 2.12, V= 616, p= 1, r=0).
As for the face-to-face, while there was an apparent decrease in formal cognitive assessments carried out before March to the last two months on average, this difference was not statistically significant (MD= 27.47, V= 435, p= 1, r=0).

However, I did find a significant difference for percentages of the telephone assessments comparing pre- and post-March 2020 (MD= -9.74, V= 0, p<.001, r=-.42). The same was true for videoconferencing (MD= -17.74, V= -4.30, p<.001, r=-.54) where there was an increase in videoconferencing. All effect sizes for significant tests (r = -. 37 to -.56), when converted to Cohen's (d = -0.80 to -1.35) were above the minimum effect size I could detect in the sensitivity analysis with 80% power.

The same patterns of change data were present in the non-UK sample data, with a few exceptions. There were seven respondents after exclusions for incorrectly answering items 9a-c to 10a-c. The non-UK group conducted similar cognitive assessments in the average month (both had a median of 10), and similar telephone

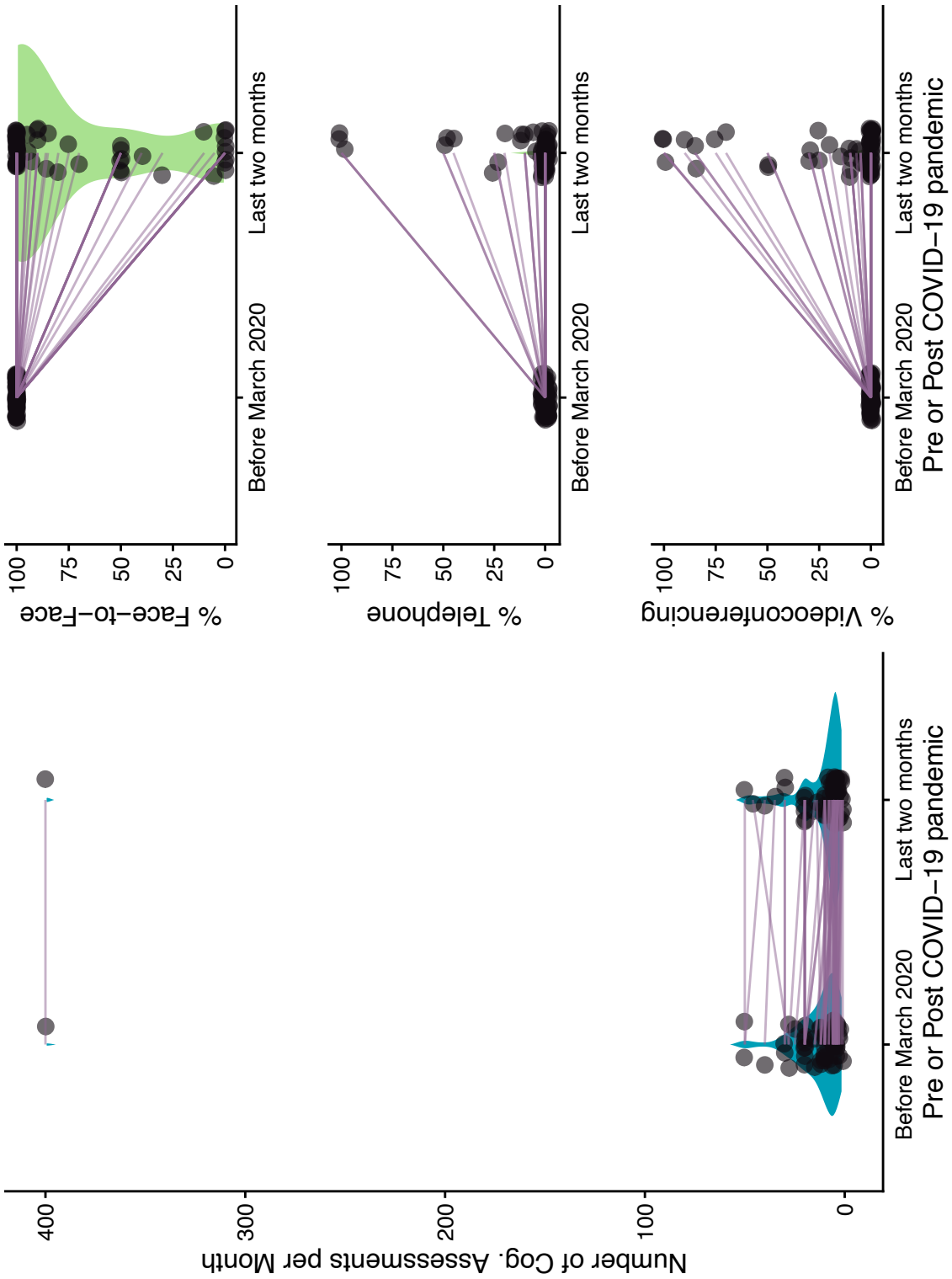


Figure 1.2: The left-side graph illustrates the number of cognitive assessments conducted in the average month, both before and during the COVID-19 pandemic, lines represent the connection for individuals' data before and during the COVID-19 pandemic. On the right-side, the percentage of assessments which were telephone or videoconferencing both before and during the COVID-19 pandemic are presented, where lines represent the connection for individuals' data before and during the COVID-19 pandemic. Figure available at https://osf.io/b6rjq/ under a CC-BY4.0 license.

and videoconferencing assessments (0% respectively) pre-March 2020. However, of the seven, only one shifted from 100% face-to-face assessments to 50% video-conferencing and face-to-face.

Cognitive Assessments Used

From the UK sample, 18 respondents gave details of specific tests they have used remotely. The most commonly reported test to remotely assess cognition pre-March 2020 was the Addenbrookes Cognitive Examination (ACE)[176] (38.89% of 18 respondents reported use), followed by the Montreal Cognitive Assessment (MoCA) [37], followed by the Delis-Kaplan Executive Function System [177] (22.22%) and the Oxford Cognitive Screen (OCS) [38] (22.22%). 16.42% of respondents reported not using the same tests during the pandemic as they did before, and eight respondents reported the tests they now use. These included the Repeatable Battery for the Assessment of Neuropsychological Status [178] (n=1), the ACE (n=2), mini-MoCA (n=1), MoCA blind (n=1), telephone MoCA (n=1), the D-KEFS (n=2), the Behavioural Assessment of Dysexecutive Function [139] (n=1), the Wechsler Adult Intelligence Scale [179] (n=2), Wechsler Memory Scale [180] (n=1), the OCS (n=1), and the Telephone Interview for Cognitive Status [91] (n=1).

In the non-UK sample 14 respondents reported tests used remotely at any time point, the most commonly used test was the MoCA (35.71%) followed by the D-KEFS (28.57%), and Wechsler Test of Adult Reading [181]. There were two respondents who changed tests since March 2020, these included starting to use the telephone Mini-Mental State Examination [48], and child specific tests. In Table 5.2 the appendix, I present all of the tests reported to be used in remote assessment. Following this analysis, data are no longer split by country.

I also investigated the OCS specifically, for aim 2, regarding its use before and after March 2020, to explore the uptake of remote assessment using the OCS. Prior to March 2020, 51.75% of all 114 respondents reported using the OCS. Users of the OCS were primarily occupational therapists (68.33%), however, other users included clinical psychologists (8.33%), medical doctors (6.67%), neuropsychologists (6.67%),

physiotherapists (3.33%), assistant psychologists (1.67), speech and language therapists (1.67%), others (3.33%) including a clinical researcher and dietician. The OCS was primarily used in acute stroke units (19.78%) and was also used by community stroke teams (16.48%), community rehabilitation units (15.38%, early supportive discharge (14.29%), neuro out-patient facilities (12.09%), mixed acute and rehab settings (10.99%), memory clinics (2.20%), and others (8.79%) including a cognition clinic, neurological in-patient (n=2), community neuro service, neurological rehabilitation (n=4), brain injury service, and geriatric clinic. The 86 respondents reported using the following groups: Stroke = 57, Traumatic brain injury = 17, Mild cognitive impairment = 9, Dementia = 2, and other = 1. Where "other" was for a brain tumor.

Since March 2020, 43.86% of respondents reported using the OCS. Of those who responded 'Yes' to using the OCS since March 2020, 18% reported using the remote assessment version. Respondents reported that they liked using OCS because it was aphasia friendly, easily adaptable to remote usage and easy to administer in comparison to other tests. The potential improvements that could be made were largely that the entire OCS should be remote rather than requiring a pack to be sent ahead of time with materials (33.33% of respondents).

Of the 58 responses on reasons they did or did not administer the remote OCS, for those who had administered it (n=9), the most common reasons reported were to assess cognition (55.55%), and integration into wider remote assessment of the professional service they operated in (22.22%). For those who did not administer the remote OCS (35 respondents), the most common reasons reported was working in in-patient facilities with limited need for remote assessment (37.14 %) as well as a lack of awareness of a remote version (22.85%).

Perceived Benefits, Challenges, and Facilitators to Remote Assessment of Cognition

I asked respondents who reported using cognitive assessments, what they experienced as benefits of remote assessment. The most commonly reported

theme emerging from the data, of 97 responses, reflected the economic, time, and environmental cost savings associated with remotely administering tests (43.30%). The second most common benefit was safety in relation to reduction of COVID-19 exposure and infection control (31.96%). This was followed by reported increased accessibility of participants attending appointments, particularly travel and rural communities were frequently mentioned (21.54%).

Respondents (n=97) reported that the key barrier to remote assessment of cognition was that they could not pick up on subtleties of behaviour or see how participants were responding to the questions/tasks (28.87%). This lack of observation made it harder to interpret the results of the assessments for some participants. The second most frequently reported barrier was that some participants could not be tested due to lack of knowledge about technology and therefore hesitance to use it (27.84%). Some respondents reported that their participants were anxious about adopting remote assessment because it was new to them. Thirdly, respondents reported concerns regarding the reliability of internet or technical issues making the connection unstable when conducting remote assessments (26.80%). Other notably frequent concerns regarded the lack of (in terms of existence or their access to) standardised and validated/normed assessments to use for remote assessments (17.50%).

Finally, when asked what could facilitate remote assessment of cognition, the most frequent response was to increase the number of standardised, validated, and normed assessments (27.84%) available for use. Other responses such as increasing access to remote assessment software (13.40%), technology (13.40%), and physically supporting the participant during administration by a family member/someone else (13.40%), were relatively equally frequent.

1.1.5 Discussion

I presented a pre-registered web-based survey study that investigated changes in the assessment of cognition before and during the COVID-19 pandemic. A sample of 114 (135 before exclusions) healthcare professionals from various clinical

settings (including primary care, in-patient settings, and specialised neurological out-patient services), working in a diverse range of roles (from physiotherapists, occupational therapists, neuropsychologists, and medical doctors), responded to the survey. This heterogeneous sample reflects the variety of professions that engage in the assessment of cognition in clinical practice. One of the key aims was to reflect responses from this clinical reality rather than limit the survey to only neuropsychologists/psychologists performing cognitive assessments. The majority of the sample were from the UK, but there were responses from across the globe. To allow meaningful interpretation of the results of certain subsection, I analysed the UK subsample separately, though the broad findings of increased remote cognitive assessments since March 2020 was evident throughout the full international sample.

Importantly, I found that the overall number of cognitive assessments used in clinical practice did not significantly differ compared to before the pandemic, which provides reassurance that this important aspect of assessing hidden deficits in neurological populations remained included in clinical services. It is testament to the ingenuity and flexibility of teams to swiftly adapt their ways of working and to keep these services running during this period of the COVID-19 pandemic.

I found that the frequency with which videoconferencing or telephone conferencing modalities were used to assess cognition increased significantly with 9.74% and 17.74% for video and telephone conferencing respectively. There were statistically more video and telephone assessments of cognition since the pandemic. This in line with what recent research has shown in other samples [168, 170]. In terms of overall remote assessment prior to the pandemic, the results did not match that of Hammers et al. [170] who found a self-reported use of video neuropsychological assessment of 11% prior to the pandemic, whereas I found that rates of remote assessment were less than 1% across the whole sample. In addition, Hammers et al. [170] found that there was no significant increase in video usage (from 11% to 15%). This may be because of differences in the respondent samples. Hammers et al. [170] reported data from over 80% neuropsychologists and psychologists, whereas I had a diverse range of roles and settings for the sample. It is highly likely that many of the occupational

therapists worked in settings where no telehealth was available, and this was only implemented following the COVID-19 restrictions. Indeed, I found that none of the sample used remote assessment pre-pandemic, and during the pandemic nearly half (45.2%) of respondents reported using videoconferencing for cognitive assessment.

I found that the most commonly used tests to assess cognition before the pandemic were the Addenbrookes Cognitive Examination [176], the Montreal Cognitive Assessment [37], the Delis-Kaplan Executive Function System [177], and the Oxford Cognitive Screen (OCS) [38]. There was limited change in tests used since the pandemic, when switching to remote assessments. This means that many of the same tests were used pre- and post-pandemic, when they were not validated in the new remote format, potentially affecting the interpretation of test performance, and making clinical decisions more complex than usual. Half of all respondents reported that they used the OCS, and this only slightly shifted down during the pandemic. The main reason that respondents did not administer remote tests was that they were working in in-patient settings where this was not required. For the remote version of the OCS, which remains to be validated in this format, a common reason for not administering the test was not knowing it existed. There is a clear need for improvement in creating and validating tests for remote usage and improved dissemination to health professionals in clinical practice of these validated modalities as soon as they become available.

With regards to the benefits, challenges, and facilitators of remote assessment of cognition, respondents reported that remote assessment had economic and time saving implications, they were COVID-19 secure, and that they increased accessibility for hard-to-reach participants who lived far away or did not travel. However, in terms of challenges faced, respondents felt that they could not get enough information from simple observation of behavioural responses which would allude to subtle impairments or different conclusions, and that the clients themselves were hesitant to use technology generally. While less frequent, an important issue that arose was the lack of standardised and validated tool for the remote assessment of cognition.

Limitations and Future Research

In order to advance the evidence base on the use of remote cognitive assessments, I intended to collect data from a web-based survey from a global sample. However, the vast majority of the recruited sample were from the UK despite attempts for a wider distribution through social media and international professional bodies such as the International Neuropsychological Society. As such I could not draw firm conclusions about the global use of remote cognitive assessments, or examine any differences depending on geographical area and severity of the pandemic. Instead, I focused statistical analyses on the UK only sample and made comparisons to the smaller non-UK sample. However, there were only very small sample differences suggesting that the observed rates in the UK may be reflective of a more global picture in regard to shifts in modality of formal cognitive assessments.

I used word of mouth, social media, websites, email and direct communication with UK based professional groups and this potentially restricted the sample. Moreover, my recruitment avenues likely resulted in a selection bias of typically engaging respondents who were motivated to take part in research rather than reflecting the full breadth of allied health professionals conducting cognitive screening. The sampling outcome was similar to Hammers et al. [170], who recruited mostly USA based respondents with smaller proportions from elsewhere in the globe, and as such is not a unique problem of this paper. In future, in order to garner a truly global sample, greater collaboration between international researchers and the inclusion of more international societies or other relevant professional bodies based in other countries could aid in gaining better sample representation.

However, I did gain a large enough sample for statistical power, and the UK analysis replicated even in the smaller more internationally diverse samples in this paper. Moreover, I aimed to capture the update and shift in the first critical phase of the pandemic, and this will be invaluable to shape future investigations.

Finally, I had a number of participants who incorrectly answered percentage based items (item 9 – 10c) because their answers for the percentage of formal cognitive assessments conducted either face-to-face or remotely did not sum to

100% or to the total number of tests they ran. In retrospect, a response validation where responses must sum to 100% and a notification of error where they do not, would have prevented data loss for those items.

Further research may also want to investigate whether there was a change in the available medical resources or staff, which may have impacted on the criteria for proceeding cognitive evaluation itself. Especially in places with very severe waves and high mortality due to the pandemic, this could, at least temporarily, have impacted the frequency of cognitive assessments, remote or otherwise.

Based on the outcomes of this study, and current research, there is an imperative to generate validated, standardised, and feasible tools to assess participants remotely either through videoconferencing, telephone conferencing, or through both modalities. This is instrumental in enabling clinicians to have greater confidence in the conclusions given by remote assessments and have clearer normative data to compare performance to.

1.1.6 Conclusion

The current study investigated shifts in the formal assessment of cognition before and during the COVID-19 pandemic in a wide range of allied health professionals involved in assessing cognition. I found that the number of cognitive assessments carried out, and the number of face-to-face assessments were not different before or during the pandemic. However, there was a large increase in the use of video and telephone conferencing modalities in assessing cognition. The study highlights the immediate need for standardised and validated tests which can give greater confidence for clinicians administering them remotely and for participants to partake in.

1.2 Introducing the Tele-OCS: A validated remotely administered version of The Oxford Cognitive Screen

Published with open peer review in Health Open Research as **Webb, S. S.**, *Carrick, C. Kusec, A. & Demeyere, N. (2023). Introducing the Tele-OCS: A validated remotely administered version of The Oxford Cognitive Screen, Health Open Research, doi: 10.12688/healthopenres.13291.1*

1.2.1 Abstract

Background Remote cognitive assessments are increasingly used with the rising popularity of teleneuropsychology. Here, I evaluated the performance of the remotely administered Oxford Cognitive Screen (Tele-OCS) compared to in-person administration in adult stroke survivors. **Methods** 40 stroke survivors (M age = 69.30, SD = 10.44; sex = 30% female) completed in-person and remote versions of the OCS on average 30 days apart, with different trained examiners. The order of administration was counterbalanced. Cohen's d estimates were used to compare performance between modalities. **Results** I found that the proportion of OCS subtasks impaired did not differ across modalities (d = <.001). With regards to raw subtask scores, only the picture naming subtask and executive score from the trail making subtask were found to be statistically different across modalities, though raw differences were minimal (<1 point difference on average). These statistical differences did not affect impairment classifications. **Conclusions** The Tele-OCS classified cognitive impairments in a comparable way to the in-person version. The validation of the Tele-OCS allows for remote assessment to increase accessibility and pragmatically aid in addressing the clinical need for stroke-specific cognitive screening in a wider population.

1.2.2 Introduction

The current study aimed to validate a remote version of the Oxford Cognitive Screen [38] that can be administered via telephone or videoconferencing. Interim provisional guidance for administration of a remote version of the OCS was released in May 2020 [182], following frequent requests to the authors. As the section 1.1 showed, the OCS was being used in 2020 through remote administration, however, no formal validation of a remote delivery format of OCS had been conducted until the current study.

The purpose of this study was to compare classification of impairment on the Oxford Cognitive Screen when administered in-person versus remotely (via telephone or videoconferencing) at least 6-months post-stroke. 6-months post-stroke was used due to current recruitment protocols which enable recruitment of stroke survivors at hospital and then follow up at 6-months post-stroke. The pre-registered hypotheses (https://osf.io/ev53t) were as follows and any deviations from the pre-registration are reported transparently: 1. Each OCS subtask continious performance score will have moderate to strong associations across in-person and remote versions (ICC >= .50); 2. The sensitivity and specificity values of the in-person and remote version of the OCS will be approximately equitable.

1.2.3 Methods

I report how I determined my sample size, all data exclusions (if any), all manipulations, and all measures in the study (21 word solution for transparency by Simmons et al. [171]). Approval for the study was granted by the Medical Sciences Interdivisional Research Ethics Committee (REC REF: R58224/RE001).

Participant Identification

Eligibility criteria were 1) history of confirmed stroke; 2) 18 years of age or older at time of stroke; 3) capacity to consent to research; 4) able to remain alert for at least 30 minutes, and 5) spoke and understood sufficient English. Participants were excluded if they had hearing, language, or visual impairments (not including visual neglect) that would not allow for remote or in-person assessment outside

of reasonable adjustments. I acknowledge the substantial selection bias of this sample, in that being able to concentrate for 30 minutes and have capacity to consent immediately exclude those more severely impaired by their stroke. The justification for such selection criteria is to remain standardised to our recruitment criteria across all studies run in group to be able to compare and contrast research.

Study Design and Sample Size

I used a prospective cross-sectional within-subjects design. I recruited community-dwelling chronic stroke survivors in the UK between November 2021 to August 2022 who had previously taken part in the lab research. Sample size was pre-determined at a minimum of 30 participants to detect an intraclass correlation coefficient (ICC) of greater than 0.50, with 90% power, and an alpha of 0.05.

1.2.4 Changes from pre-registration

This project was originally pre-registered at the start of data collection (3 participants had taken part). The preregistered analysis plan was inspired by parallel forms analysis in the original OCS publication [38], whereby ICCs were conducted between parallel versions A and B. Once data was collected and analysed, it became clear that this was not the best approach for the observed data. The small ranges of observed values in the subtasks meant the continuous performance scores could not be analysed using ICCs.

Instead, in line with the clinical relevance of cognitive screening, I used both binary impairment classifications for subtasks and overall performance to determine if deviations between in-person and remote assessments would change interpretation of overall performance, and t-test equivillance tests.

I re-ran the analysis after collecting further participants (n = 9) who completed only OCS version A both remotely and in-person due to the timeline in which I changed the analysis plan. I included these participants in the final dataset. Power to detect effects was virtually unchanged when adapting from ICCs to equivillance t-tests and impairment classification data.

1.2.5 The Oxford Cognitive Screen (OCS)

The OCS was previously described in the thesis, but see Demeyere et al. [38].

1.2.6 The Tele- Oxford Cognitive Screen

The remote version of the OCS is an adaptation of version A of the OCS. Administration of the Tele-OCS can be conducted via telephone or video-call. Before the appointment, participants are sent a single testing booklet through the post containing the necessary stimuli to complete the OCS remotely. Participants are instructed not to open the pack before the phone or videocall appointment and are asked to ensure they sit at a table for the duration of the session and have a pen or pencil to hand. The assessor keeps a copy of the remotely sent materials in front of them to jointly follow along with the participant through each subtask. During the session, participants are given specific instructions (provided in the user manual for the examiner) on how to complete each subtask of the Tele-OCS, including the relevant page to turn to in their testing booklet, when to pick up and put down their pen or pencil, and which way to position their booklet on the table. If completing the Tele-OCS via telephone call, participants are instructed to put the telephone on speaker mode, allowing participants freedom to use both hands to complete the subtasks. Upon completion of the assessment session, participants were asked to place the testing booklet into a pre-paid addressed envelope and return it to the assessors via the post. Where necessary, facilitation of page turning and reinforcing assessor instructions was supported by a carer or family member.

Several differences exist between the in-person and remote version of the OCS. Given that the praxis and visual field subtasks require the participant to be in full view of the assessor, these are not assessed as part of the Tele-OCS. To assess constructional praxis, the Tele-OCS instead contains a figure-copy task, requiring participants to copy, and immediately recall, a complex figure. The figure was taken from the OCS-Plus assessment, under the same copyright as the OCS [38]. The spatial attention task was adapted for remote administration by giving two

practice rounds, instead of one, to complete before the main task. This was put in place to allow the examiner to gauge whether the participant has understood the instructions, without being able to see the practice trials. On the first round, participants practice a short version of the spatial attention subtask, where they are presented with six line-drawings of hearts, three of which contain a gap on either the left or right-hand side. Participants are instructed to only cross out the complete hearts on the page. On the second round, participants verbally state which of these hearts have a gap and which ones are complete. This additional practice round allows the assessor to evaluate whether the participant has understood the instructions before they proceed to the main task.

Finally, adaptations were made to the remote version of the OCS trail making task. When this subtask is administered in-person, participants are shown a completed example by the assessor which they can refer to throughout the duration of the task. As this is not possible through remote administration, instructions have been printed at the top of the practice task and the main task in the remote testing booklet. The materials for the Tele-OCS and the manual for administration are freely available for publicly funded clinical and research use in the same way as the in-person OCS. The Tele-OCS is licensed through Oxford University Innovations Health Outcomes (https://innovation.ox.ac.uk/outcome-measures/the-oxford-cognitive-screen-ocs/).

I timed the duration of the Tele-OCS administration for a portion of participants only (n = 18). I additionally recorded any pragmatic information regarding the optimal assessment of the Tele-OCS. This involved recording unsolicited spontaneous feedback from participants on the Tele-OCS after administration.

1.2.7 Data analysis

First, I summarized the administration time of participants and comments given by participants about the remote OCS. Cut offs for impairment per OCS subtask were created using the original OCS normative cut offs from healthy ageing data [38]. Then I visually depicted proportion of impaired subtasks per modality and

impairment classification per sub-task for both modalities. I further explored continuous performance scores of participants who were classed as unimpaired on the in-person OCS but impaired on the remote OCS, and vice versa. I compared raw continuous performance scores for statistically significant differences uses t-tests and Cohen's d effect size.

1.2.8 Results

Sample

Forty chronic stroke survivors took part in this study, with demographics of participants presented in Table 1.2. The median time post-stroke was 228 days (IQR=189.25 to 1500.50). One participant did not return their remote OCS pack, as such I was unable to generate scores for the number writing, broken hearts, or trails accuracy subtasks. I include this participant in the majority analyses where complete data were available, including proportion of subtasks and domains impaired (proportion accounts for missing subtasks).

Table 1.2: Table 1. Demographics of the participants included in the final analysis.

Characteristic	N	Values
Age (M(SD))	40 (0)%	69.3 (10.44)
Education (M(SD))	40 (0)%	13.45 (2.85)
Sex	40 (0)%	F: 30%; M: 70%;
Ethnicity	40 (0)%	Black-Carribean: 2.5%; White-British: 97.5%;
Days Since Stroke (M(SD), [range])	39 (2)%	842.21 (971) [122 to 3283]
Stroke type	40 (0)%	Subarachnoid haemorrhage: 2.5%; Intracerebral haemorrhage: 27.5%; Ischaemic: 70%;
Stroke side	40 (0)%	B: 10%; L: 45%; R: 45%;
NIHSS (Median(IQR))	37 (8)%	7 (3 to 9)

Tele-OCS overview

The time in days between in-person and remote OCS administration ranged from 6 to 72, with an average of M=27.70 days (SD=12.29). Twenty-three participants were administered the remote OCS first (57.5%). Administration of the remote

OCS took an average time of 17.17 minutes (SD=2.73, range = 14-22). The remote OCS was administered using phone to the majority of participants (n=36, 90%). Fourteen participants volunteered feedback after the remote OCS was administered. Nine participants commented that there were limited differences in their experience between in-person and remote administration. Three participants had issues with their phone connection, which broke up the session slightly, but minimally impacted administration. One person who had difficulty finding the correct pages due to a hemiplegia, required help from his partner to turn the pages. One participant required a repetition of one instruction once due to hearing difficulties. Otherwise, participant remote sessions were administered without major barriers.

Comparison of the Tele-OCS and standard OCS

I examined the proportion of subtasks impaired on the in-person versus remote OCS, and then the frequency of impairment classifications for each OCS subtask across modalities. Participants were impaired on an average of 5.71 (SD=7.79) subtasks on the in-person version and 5.71 (SD=8.59) subtasks in the remote version. In Figure 1.3 I present the Kendall rank correlation coefficient for proportion of subtasks impaired between remote and in-person OCS versions. There was a moderate significant relationship between proportion of subtasks impaired across modalities. I visually depict impairment classification distributions for all OCS scores in Figure 1.4 to illustrate the similarities / differences between modalities. Figure 1.4 shows that, in terms of count, there were a few individuals who were impaired only on the remote version of the task, but not in the in-person version. In Table 1.3, I present descriptive statistics for raw score performance on each shared OCS subtask as well as repeated-measures Cohen's d and significance of difference between each subtask.

The results show that regardless of the few individuals who were different in score, there were few statistically significant differences between raw score performance on each subtask and overall proportion of tasks impaired. I present sensitivity/specificity analysis in the Supplementary Materials section as not all

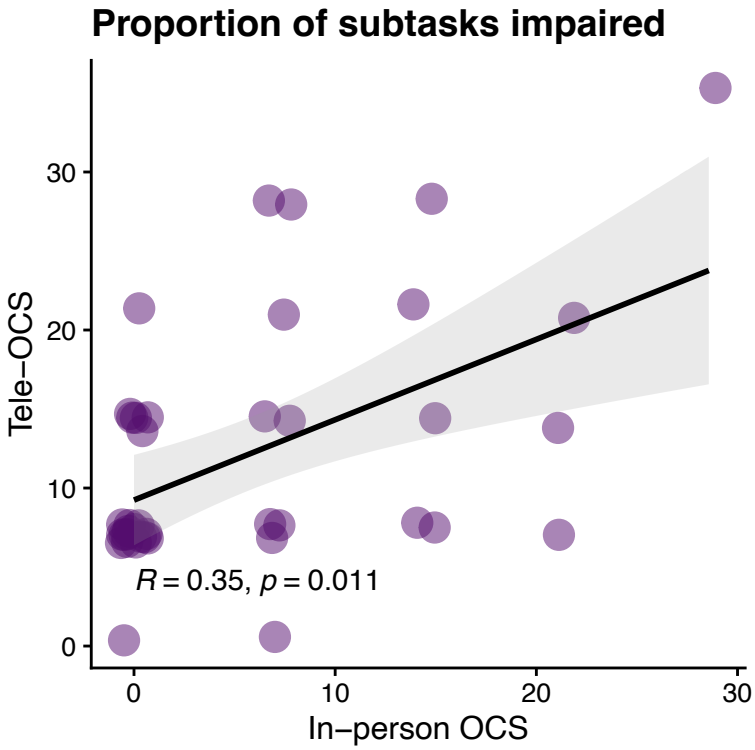


Figure 1.3: Scatter plot of the proportion of subtasks classified as impaired on the in-person OCS and remote OCS. R = Kendall rank correlation coefficient. Figure available under CC-BY 4.0 license (https://osf.io/9p7fs).

analysis could be completed due to low and differing numbers of impaired impairment classifications.

1.2.9 Discussion

I aimed to validate the use of a remote version of the Oxford Cognitive Screen (OCS) by comparing performance within-subjects on the OCS when administered in-person (OCS version B) and remotely (Tele-OCS version A), via telephone or video conferencing. Overall, the proportion of those impaired on the OCS subtasks was moderately related between in-person and remote administration. Within each of the subtasks, whilst I found small statistically significant deviations in raw scores in two subtasks (picture naming and executive score), impairment classifications on each task were minimally affected by modality.

It is possible that differences in OCS versions (A vs B) led to increased error variability in performance on OCS subtasks between modalities. This may be

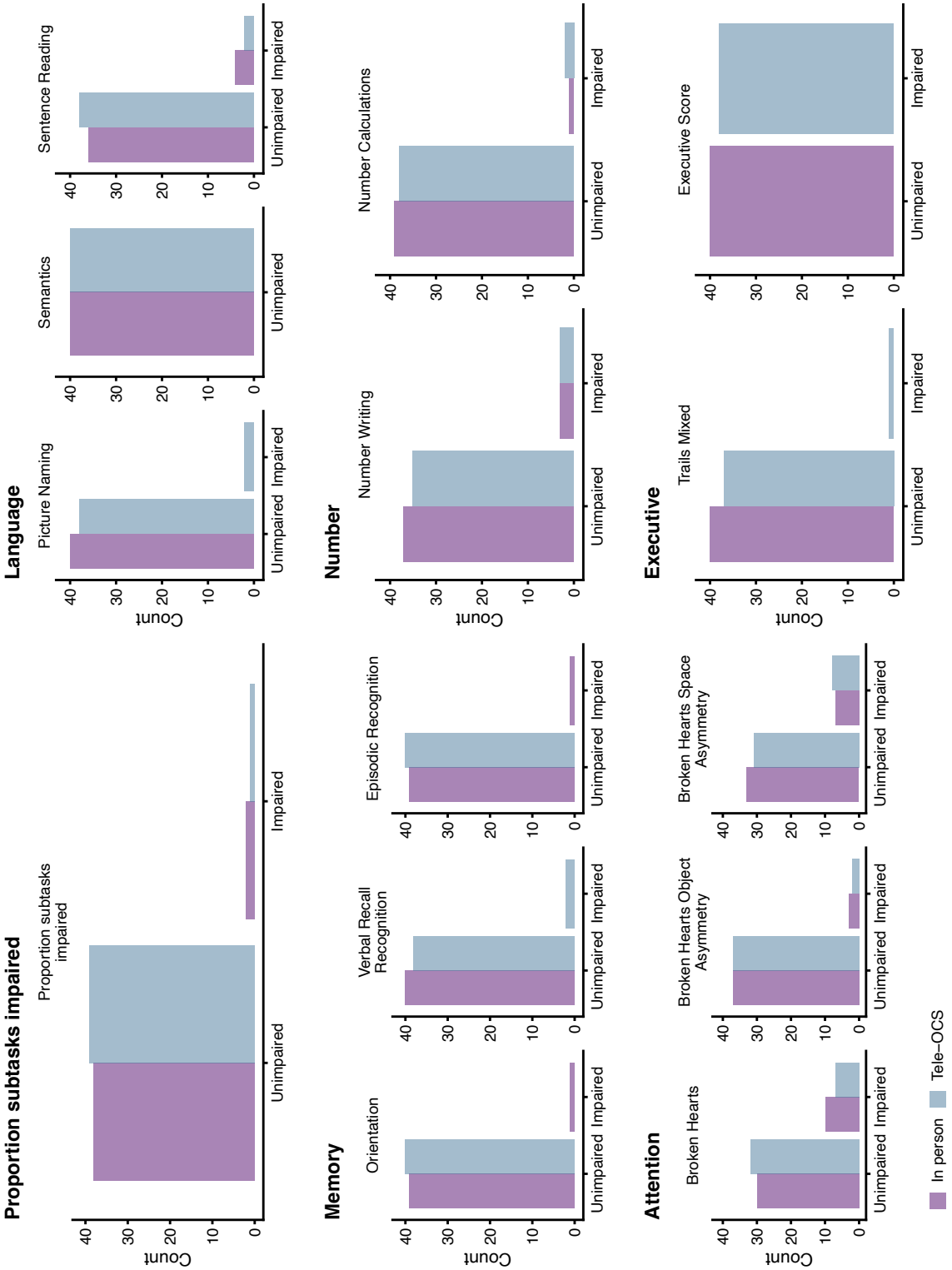


Figure 1.4: Density plots of performance on the OCS when remotely administered or administered in-person in stroke survivors. Figure available under CC-BY 4.0 license (URL).

Table 1.3: Descriptive statistics for subtasks per domain on the remote and in-person OCS and test of difference via t-test and Cohen's d.

Domain	OCS test	N	In person M(SD)	Remote M(SD)	Cohen's d
Language	Picture Naming	40	3.92 (0.27)	3.7 (0.56)	0.36*
	Semantics	40	3.05 (0.22)	3 (0)	0.23
	Sentence Reading	40	14.53 (0.82)	14.8 (0.52)	-0.31
Number	Number Writing	39	2.92 (0.27)	2.87 (0.52)	0.11
	Number Calculations	40	3.9 (0.38)	3.75 (0.54)	0.23
Memory	Orientation	40	3.98 (0.16)	4 (0)	-0.16
	Verbal Recall Recognition	40	3.95 (0.22)	3.83 (0.59)	0.24
	Episodic Recognition	40	3.88 (0.4)	3.8 (0.41)	0.13
Attention	Broken Hearts	39	43.85 (7.96)	44.67 (6.41)	-0.2
	Broken Hearts Object Asymmetry	39	0.17 (0.68)	0 (0.83)	0.24
	Broken Hearts Space Asymmetry	39	0.5 (2.65)	0.67 (2.74)	-0.05
Executive	Trails Mixed	39	12.45 (1.15)	11.74 (2.11)	0.42*
	Executive Score	39	-0.68 (0.86)	0.1 (2.02)	-0.46*
Subtasks impaired	SubtasksImpaired Impaired (%)	40	5.71 (7.79)	12.14 (8.12)	-0.8*

Note. '*' for t-test p=value is <.05.

particularly the case for the picture naming task, where two participants were impaired on the remote OCS (based on version A), but not on the in-person OCS (based on version B). There are small differences between version A and B of the picture-naming task, irrespective of modality. For example, participants are more likely to make errors on version A, by calling the hippo a 'rhino' and calling the slices of melon, 'fruit' (see OCS control normative data, Demeyere et al., [38]; in OSF project ttps://osf.io/auwjp). For the executive score, although scores between modalities were statistically different, impairment ranges were nearly identical, with only one participant classed as impaired on the remote version and unimpaired on the in-person version. Regarding the original psychometric validation of the in-person OCS, the variabiliy in versions is illustrated with diverse intraclass correlation coefficients [38]. Note, the same issues here apply also to Demeyere et al. [38], that is, small range items reduced variance which preclude accurate intraclass correlation coefficients. Demeyere et al. [38] did not conduct alternative analyses to compare versions which would cater for reduced variance.

The results align with other validation studies of remote versions of commonly used cognitive screening tools. In a longitudinal study comparing performance on telephone versus in-person administration of a cognitive test battery of multiple domains (attention, verbal learning and memory, verbal fluency, executive function, working memory and global cognitive functioning), statistical differences were found between in-person and remote assessment scores, with the most reliable test-retest data found using the same modality [183]. Despite these small differences, remote and in-person administrations yielded equivalent interpretations in most cases, with the authors concluding that this remotely assessed battery was comparable to its in-person version. Katz et al. [90] compared performance on the MoCA and Tele-MoCA, finding small but non-meaningful differences in performance between modalities (e.g., non-significant equivalence tests). Overall, prior research, along with the results of the present investigation, suggest that slight differences between remote and in-person versions of the same cognitive test are to be expected. It is possible that continuous performance scores, when delivered remotely, may be more likely to fluctuate, possibly due to technical problems, lack of examiner presence to guide correct instructions (e.g., for the trail-making task), and hearing or language difficulties that become exacerbated during remote cognitive assessment. However, overall differences are minor, and should not be a barrier to administering cognitive tests in remote format nor to overall interpretation of participant results. The Tele-OCS thus provides a highly comparable assessment of cognitive functioning.

There are several stroke-specific considerations concerning remote cognitive assessments. For instance, motor impairments may affect stroke survivors' ability to complete cognitive tests in remote format. One participant in the present study remarked that turning pages in the remote OCS booklet was difficult due to their hemiplegia. Motor difficulties such as these should be taken into consideration by clinicians when assessing stroke survivors remotely, so that best-practice may be established. Additionally, visuospatial neglect, a common post-stroke impairment, may impact remote cognitive test performance. Participants should be reminded to keep their remote testing booklet as close to their mid-line as possible so that

cognitive performance in other domains is not impacted by neglect symptoms. The in-person and remote OCS are designed to be spatial neglect friendly, with stimuli presented centrally on all pages. Involvement of a family member or carer at home when conducting a remote assessment may be beneficial for participants that require additional support where in person assessment is not possible.

Remote cognitive assessment has multiple benefits. Delivering cognitive assessments in remote format can increase the accessibility of these assessments to a wider range of individuals [87]. Similarly, remote cognitive assessments can reduce participant burden for those participating in research studies, given that individuals do not have to be physically present to participate. Those who would not usually participate in research may be encouraged to do so if the modality of assessment is remote. It should be noted that the present investigation was the first to validate a stroke-specific cognitive screening tool for remote use. Given that commonly used remote assessments, such as the MoCA and Tele-MoCA, do not usually assess stroke-specific cognitive impairments (e.g., unilateral spatial neglect and aphasia; Laska et al., [184]; Moore et al., [185]), the remote OCS can improve the accessibility of adequate cognitive screening to stroke survivors during their recovery, increasing the likelihood of appropriate cognitive review if necessary.

Limitations

Although my study sample was statistically powered, the sample size was small and relatively homogeneous. For example, 97.5% of the sample was White-British, and most sustained a moderate stroke. Furthermore, given that many participants were in the chronic stage of stroke, they were often classified as unimpaired across subtasks in both modalities, limiting the interpretation for stroke survivors with more severe impairments. This may impact the generalisability of the present results to the wider stroke population. Future validation studies of the remote OCS may benefit from using larger, more heterogeneous samples, including those who have sustained more severe strokes.

A further limitation of the present study is the reliance on normative data from the in-person OCS for impairment classifications. It is not necessarily true that the normative cut offs will be identical across modaliites. In person cut offs here were chosen pragmatically as these were the only available. There is currently no normative data for performance on the Tele-OCS for comparison of scores. Future research, given the slight discrepancy between results should generate new normative data, even at least to validate the use of the in-person norms for the Tele-OCS.

Implications for Research and Clinical Practice

The present investigation provides insight into how to best administer the Tele-OCS. Sending a paper pack in the post allows for clear presentation of instructions and stimuli while carrying out remote OCS assessment. However, there are also several drawbacks to this method; participants must be instructed not to look at the testing pack ahead of the testing session and it can be time-consuming (and potentially costly to NHS services) to send the pack via post. However, use of a paper testing pack allows stroke survivors who do not have access to videoconferencing to carry out the testing session over the telephone, without having to travel to a research or clinic site which incurs its own cost. In the sample, the majority of participants opted to complete the remote OCS over the phone, rather than using videoconferencing. This may be because they are less familiar with, or do not have access to the required technology.

I encourage health professionals to take these factors into consideration when choosing to use the Tele-OCS. Clinicians who are new to remote assessment are encouraged to carry out practice administrations with the remote OCS before using it in a clinical context.

Conclusion

The OCS, when administered remotely, is a potentially valid method of screening for cognitive impairment among stroke survivors. Clinicians should use existing

published normative data to interpret impairment on OCS sub-tests when delivered remotely until new norms are generated.

2

Screening for subtle cognitive deficits

2.1 Validation of the UK English Oxford Cognitive Screen-Plus in Sub-Acute and Chronic Stroke Survivors

This section is partially altered from the peer-reviewed publication: *Demeyere, N., Haupt, M.,* ***Webb, S.S****. et al. Introducing the tablet-based Oxford Cognitive Screen-Plus (OCS-Plus) as an assessment tool for subtle cognitive impairments. Sci Rep 11, 8000 (2021).* https://doi.org/10.1038/s41598-021-87287-8

2.1.1 Abstract

Introduction Stroke survivors are routinely screened for cognitive impairment with tools that often fail to detect subtle impairments. The Oxford Cognitive Screen-Plus (OCS-Plus) is a brief tablet-based screen designed to detect subtle post-stroke cognitive impairments. I examined its psychometric properties in two UK English-speaking stroke cohorts (subacute: <3 months post-stroke, chronic: >6 months post-stroke) cross-sectionally. **Patients and Methods** This study included 347 stroke survivors (mean age = 73 years; mean education = 13 years; 43.06% female; 74.42% ischaemic stroke). The OCS-Plus was completed by 181 sub-acute stroke survivors and 166 chronic stroke survivors. All participants also

completed the Oxford Cognitive Screen (OCS) and a subset completed the Montreal Cognitive Assessment (MoCA) and further neuropsychological tests. **Results** First, convergent construct validity of OCS-Plus tasks to task-matched standardized neuropsychological tests was confirmed (r >.30). Second, I evaluated divergent construct validity of all OCS-Plus subtasks (r <.19). Third, I reported the sensitivity and specificity of each OCS-Plus subtask compared to neuropsychological test performance. Fourth, I found that OCS-Plus detected cognitive impairments in a large proportion of those classed as unimpaired on MoCA (100%) and OCS (98.50%). **Discussion and Conclusion** The OCS-Plus provides a valid screening tool for sensitive detection of subtle cognitive impairment in stroke patients. Indeed, the OCS-Plus detected subtle cognitive impairment at a similar level to validated neuropsychological assessments and exceeded detection of cognitive impairment compared to standard clinical screening tools.

2.1.2 Introduction

The OCS-Plus is a computer-tablet based cognitive screening tool that has been developed to screen for subtle post-stroke cognitive impairment (in particular, impairments in executive attention and memory)[50] using a reflective measurement model [186]. The OCS-Plus is an extension of the OCS, which is routinely used in clinical practice to screen for early stroke-specific cognitive deficits. Like the OCS, the OCS-Plus was designed to minimize language demands, cultural confounds, and examiner bias. Administration is standardised via a platform independent app and takes approximately 25 minutes to administer. Following administration, automated performance reports are provided, based on matched age-specific cut offs. The OCS-Plus has been standardised and normed in UK and German populations and preliminary psychometric validation has been completed in healthy ageing adults [50]. OCS-Plus performance meaningfully varies with socio-economic factors and age, demonstrating its sensitivity as a cognitive screening tool [187].

Here, I present psychometric validation evidence for the OCS-Plus in a sub-acute (<3 months post-stroke) and chronic (>= 6 months post-stroke) stroke cohort. This

is the first study to investigate the validity of the OCS-Plus in a clinical stroke cohort. First, I assessed construct validity by comparing OCS-Plus task performance to sub-task matched validated standardised neuropsychological tests. Second, I evaluated sensitivity of detecting cognitive impairment with OCS-Plus compared to clinically used first-line screening tools (OCS and MoCA) and compared to standardised neuropsychological tests.

2.1.3 Methods

I report how I determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (21 word solution for transparency by Simmons et al. [171]). Researchers were trained in administration of assessments included in this study, including training and certification for the MoCA administration. The manuscript adheres to the STROBE cohort study checklist [188].

Standard Protocol Approvals, Registrations, and Patient Consents

Participants were recruited for the OCS-Recovery study, under ethical approval of the South Central - Oxford C Research Ethics Committee (Ethics Ref: 18/SC/0550l; IRAS Ref: 248483; Protocol number PID 13803). Participants were recruited at the John Radcliffe acute stroke unit and the Oxfordshire Stroke Rehabilitation Unit. Participants were eligible for inclusion if they were >=18 years of age, able to sufficiently comprehend English and had a suspected/confirmed stroke. Participants were excluded if they could not provide informed consent, were too unwell to concentrate for approx. 30 mins (as judged by the multidisciplinary team), or had severe sensory impairments which meant they could not sufficiently see the stimuli or hear the instructions. All participants provided written informed consent. The construct validation section of this study was pre-registered (pre-registration osf.io/t8zug).

2.1.4 Participants

This investigation included all OCS-Recovery Study participants who had completed the OCS-Plus in either the sub-acute (n = 181) or chronic (n = 166) stage post-stroke. Sub-acute testing was conducted in an inpatient clinical setting (acute stroke unit or stroke rehabilitation unit) between 2015-2016 and 2020-2022, and chronic testing was conducted in stroke survivors' homes between 2015-2020. Sub-acute and chronic participants do not overlap. Table 2.1 reports demographic information of the sample.

Table 2.1: Clinical and demographics details for subacute (<3 months post-strok) and chronic (>=6 months post-strok) stroke samples.

	All		Subacute		Chronic	
Characteristic	N	Sample	N	Sample	N	Sample
Age (M(SD))	346 (0)%	72.85 (13.36)	181 (0)%	71.77 (13.68)	165 (1)%	74.04 (12.93)
Education (M(SD))	255 (27)%	13 (3.46)	163 (10)%	13.17 (3.44)	92 (45)%	12.72 (3.48)
Handedness	324 (7)%	A: 0.93%; L: 9.26%; R: 89.81%;	158 (13)%	R: 100%;	166 (0)%	A: 1.81%; L: 18.07%; R: 80.12%;
Sex	347 (0)%	F: 42.94%; M: 57.06%;	181 (0)%	M: 100%;	166 (0)%	M: 10.24%; F: 89.76%;
Ethnicity	347 (0)%	Asian-Other: 0.29%; Black African: 0.29%; Chinese: 0.29%; Mixed White Black Caribbean: 0.29%; Senegal-Wolof: 0.29%; White-Bulgarian: 0.29%; White-Portuguese: 0.29%; Black-Caribbean: 0.58%; Other-Asian: 0.58%; White-English: 0.58%; Other: 0.86%; White-Spanish: 0.86%; White-Other: 3.46%; White-Unknown: 10.95%; White-British: 80.12%;	181 (0)%	White-Bulgarian: 0.55%; White-Portuguese: 0.55%; White-English: 1.1%; White-Spanish: 1.66%; White-Other: 6.63%; White-Unknown: 20.99%; White-British: 68.51%;	166 (0)%	Asian-Other: 0.6%; Black African: 0.6%; Chinese: 0.6%; Mixed White Black Caribbean: 0.6%; Senegal-Wolof: 0.6%; Black-Caribbean: 1.2%; Other-Asian: 1.2%; Other: 1.81%; White-British: 92.77%;
Days Since Stroke (M(SD))	340 (2)%	102.85 (94.21)	180 (1)%	19.37 (16.37)	160 (4)%	196.76 (43.09)
Stroke type	301 (13)%	Other: 0.66%; SAH: 1%; CVA: 3.32%; TIA: 3.32%; ICH: 17.28%; Ischaemic: 74.42%;	169 (7)%	Other: 1.18%; CVA: 1.78%; SAH: 1.78%; ICH: 14.2%; Ischaemic: 81.07%;	132 (20)%	CVA: 5.3%; TIA: 7.58%; ICH: 21.21%; Ischaemic: 65.91%;
Stroke side	304 (12)%	B: 8.55%; L: 40.13%; R: 51.32%;	171 (6)%	B: 9.94%; L: 35.09%; R: 54.97%;	133 (20)%	B: 6.77%; L: 46.62%; R: 46.62%;
Stroke Severity	267 (23)%	7.49 (5.8)	123 (32)%	7.44 (5.13)	144 (13)%	7.54 (6.33)

An a priori power analysis was conducted to examine construct validity at a correlation of $r>=.30$ for convergent correlations, with 80% power, one-sided, with an alpha level of .05. This indicated a minimum sample requirement of 66 participants. No a priori power analysis was conducted for sensitivity analyses and all available data was used. A benchmark of $r>=.30$ was used as it is a benchmark in psychometric validation evidence for convergence [129].

Cognitve Data

All participants completed the computer tablet based OCS-Plus [50]. The OCS-Plus includes 10 subtasks and scores 18 impairment types based on normative data. See previously published studies for further detail [50, 187, 189]. All participants also completed the Oxford Cognitive Screen (OCS) [38] as a first-line cognitive screening tool for stroke.

Further cognitive data for validation in the subacute group was collected in several sessions. In the first session, OCS-Plus was conducted, then, based on convenience and availability of both researcher and participant, up to two follow up sessions were conducted to collect further validation data with the Montreal Cognitive Assessment (MoCA) [37] and the neuropsychological test battery. Order of administration was dependent on the examiner with one DClin student first administering MoCA and myself first administering the test battery. Figure 2.1 depicts a flowchart and participant numbers completing each of the follow-up sessions. The chronic stroke survivors only completed a single session.

OCS-Plus and OCS All chronic stroke survivors completed both OCS and OCS-Plus in a single session (n=166). The OCS-Plus research version was administered via a MATLAB [190] executable file on a Windows Surface Pro tablet. Of the 181 sub-acute participants who completed OCS-Plus, 178 had completed OCS as part of standard clinical screening practice within 30 days prior to OCS-Plus and this data was collected via the patient records. Thirty days was chosen as a time based cut off for OCS to OCS-Plus comparisons to maximise the inclusion of the OCS data.

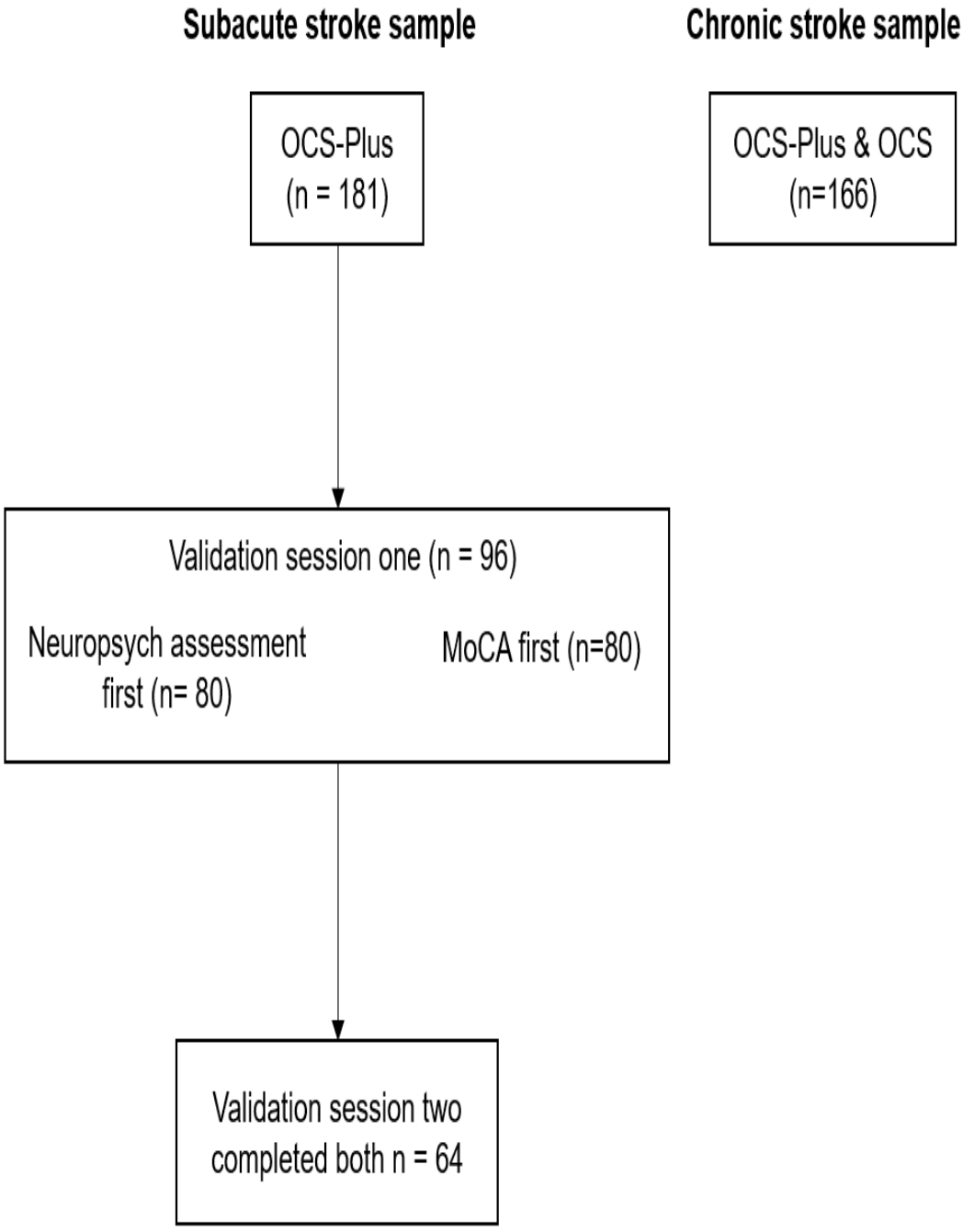


Figure 2.1: Flow diagram of the data collection streams across subacute (<3 months post-stroke) and chronic (>=6 months post-stroke) for the convergent validity portion of the current study.

Validation tests MoCA was completed as a comparison to the OCS-Plus specifically in 80 participants. Domain-specific matched neuropsychological tests were completed by 80 sub-acute participants. Each neuropsychological comparison test was chosen to be as matched in terms of format of task and administration and construct measure as possible. These included: the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) figure copy and figure recall tasks which validly measures visuospatial perception/construction and memory [178] to compare to OCS-Plus figure drawing; Behavioural Inattention Test (BIT): Star Cancellation which validly measures visuospatial neglect [101] to compare to OCS-Plus cancellation; Brixton Spatial Anticipation (Brixton) test which validly and reliably measures executive functioning [191–193] to compare to OCS-Plus rule finding; and the Cognitive Linguistic Quick Test: Symbol Trails (CLQT) which validly measures both visuospatial scanning and executive functioning [194] to compare to the OCS-Plus trail making test. Corsi block test is a computer-tablet form of the Corsi block working memory test [195] to compare to OCS-Plus invisible cancellation. These tests were scored using established cut-offs (see Table 6.1 in the appendix for details). License requirements for use of the MoCA were met.

Data analysis

First, task-specific OCS-Plus convergent and divergent validity was examined against task matched validated neuropsychological tests in a pre-registered Spearman's Rho correlation analyses. I aimed for correlations >.30 to demarcate convergence [129] but also cautiously interpret correlations >.19 (as seen in previous OCS-Plus validation study; Demeyere et al. [50]).

Second, I investigated OCS-Plus impairment incidence and impairment sensitivity and specificity. Task-by-task OCS-Plus impairment incidence was calculated as the number of stroke survivors classified as impaired on each task (compared to age-specific normative data [50]), divided by the number of stroke survivors who completed that task. Impairment sensitivity on OCS-Plus tasks was examined versus that of first line cognitive screening tools (MoCA and OCS) using published

cut-offs (see Table 6.1 in the appendix). To further investigate the sensitivity of OCS-Plus relative to gold standard neuropsychological assessment, OCS-Plus sensitivity was compared to impairment detected on neuropsychological tests described earlier. In addition, to determine whether OCS-Plus indeed detects subtle cognitive impairments, I calculated the proportion of stroke survivors classified as unimpaired using established cut-offs on clinically used tools (OCS and MoCA) who showed cognitive impairment on the OCS-Plus. Importantly, an assumption was made that following a recent stroke, cognition is expected to be affected in comparison to a healthy ageing normative group.

No missing data for validation analyses were imputed, each analysis was only conducted on those who had complete data for each analysis, degrees of freedom or absolute N are reported per analysis transparently.

Analysis Software and scripts

All data wrangling was completed in Matlab as the executable file for the research OCS-Plus produced MatLab structures for performance data. All statistical analysis and some data sorting was computed in R studio [174] version 4.2.2. We used the following packages for the production of the RMarkdown manuscript and analysis: *bookdown* version 0.32 [196], *yardstick* version 1.2.0 [197], *readxl* version [198], *pROC* version 1.18.0 [199], *knitr* version 1.42 [200], *ggplot2* 3.4.0 [201], *magick* 2.7.5 [202], *webshot* 0.5.4 [203], and *kableExtra* [204]. The underlying collated data and the code used for all analyses/manuscripts reported are openly available through the Open Science Framework (`https://osf.io/t8zug/`) with a CC-BY 4.0 license.

2.1.5 Results

Nine participants took greater than one session to complete the OCS-Plus due to fatigue or interruptions. Including only those who completed the OCS-Plus in one day, for time taken to complete the OCS-Plus, the chronic sample took on a median of 24minutes 24 seconds and the subacute sample took 21 minutes 0 seconds.

Missing data on specific tasks could be due to motor, visual or perceptual impairments that could not be compensated for, or due to ward-based interruptions where rehabilitation was prioritised. I present the reasons for non-completion of OCS-Plus subtasks in Table 2.2.

Table 2.2: Inclusion and reasons for not Testing on all Subtests of the OCS-Plus for chronic and some subacute stroke survivors.

Subtest	% Not complete	% Visual	% Motor	% Language	% Ran out of time	% Interruptions	% Fatigue	% Other	% Technical
Picture naming	1.01	0.51	0.00	0.00	0.00	0.00	0.00	0.51	0.00
Semantics	1.01	0.51	0.00	0.00	0.00	0.00	0.00	0.51	0.00
Orientation	3.54	0.00	0.00	1.01	0.00	0.00	1.01	1.01	0.00
Encoding	2.02	0.00	0.00	1.52	0.00	0.00	0.00	0.51	0.00
Trails	4.04	1.01	0.51	0.51	0.00	0.00	1.52	0.51	0.00
Delayed Recall plus Recognition and Episodic Recognition	5.56	0.51	0.00	1.01	0.00	0.00	2.53	1.01	0.51
Rule Finding	26.77	1.52	3.54	0.51	0.51	0.00	3.03	16.67	0.00
Figure Copy	20.71	1.01	0.00	0.51	3.54	0.51	6.57	7.07	0.51
Cancellation	7.58	1.52	0.51	0.51	2.02	0.51	1.52	0.51	0.51

Note. Condition of testing data were only available for 138 chronic stroke survivors and a subset of 60 subacute stroke survivors, due to differences in data collection protocols.

Construct validity

First, I examined OCS-Plus task performance and construct validity. Figure 2.2 presents key density plots showing OCS-Plus performance distributions for healthy controls, sub-acute, and chronic stroke survivors. The figure shows that some tasks are very specific regarding identifying who does not have an impairment, while others are very sensitive to detecting impairment, this is in compliment to later sensitivity/specificity analysis. The stroke cohorts were matched for age and education. These plots show that healthy controls perform better than chronic stroke survivors on all tasks, who in turn performed better than sub-acute stroke survivors (See appendix Figure 6.1 for all density plots). Table 2.3 presents correlations between OCS-Plus tasks and matched neuropsychological tests. All OCS-Plus tasks showed convergent validity against neuropsychological tests. All OCS-Plus tasks showed divergent validity, except the OCS-Plus language and orientation tasks, which also correlated with visuo-spatial tests.

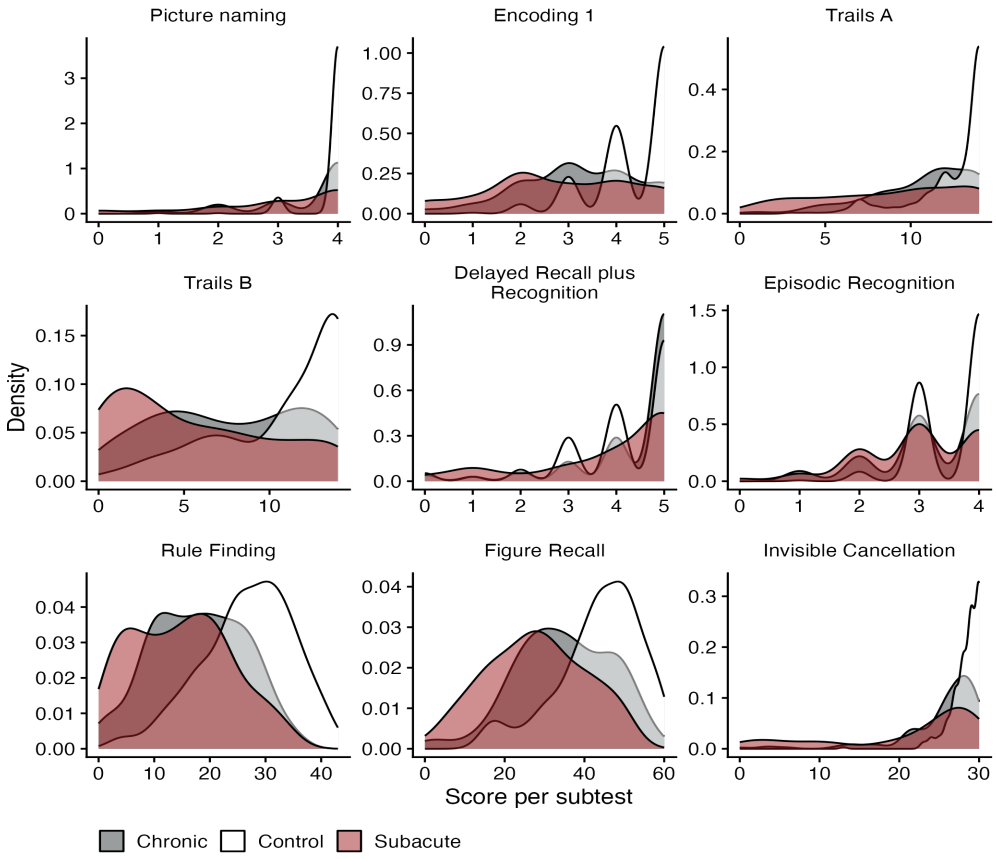


Figure 2.2: Group performance density plots per OCS-Plus task for healthy ageing adults, as well as sub-acute (<3 months post-stroke) and chronic (>=6 months post-stroke) stroke samples. Figure available under CC-by 4.0 license https://osf.io/dtmke

Table 2.3: Convergent correlations between OCS-Plus tests and construct- and format-matched neuropsychological validation tests

OCS-Plus task	Convergent test	Convergent Correlation	Divergent test	Divergent Correlation
Picture Naming Accuracy	OCS Naming Accuracy	r(340)=.35, p< .001	RBANS Figure Copy Accuracy	r(73)=.33, p=.003
Picture Naming Accuracy	MoCA Naming Accuracy	r(78)=.34, p=.002		
Semantics Accuracy	OCS Semantics Accuracy	r(340)=.17, p=.002	RBANS Figure Copy Accuracy	r(73)=.33, p=.004
Orientation Accuracy	OCS Orientation Accuracy	r(335)=.38, p< .001	RBANS Figure Copy Accuracy	r(73)=.26, p=.027
Orientation Accuracy	MoCA Orientation Accuracy	r(78)=.58, p< .001		
Encoding 1 Accuracy	MoCA Encoding 1 Accuracy	r(77)=.5, p< .001	RBANS Figure Copy Accuracy	r(72)=.15, p=.207
Encoding 2 Accuracy	MoCA Encoding 2 Accuracy	r(77)=.53, p< .001	RBANS Figure Copy Accuracy	r(72)=.21, p=.076
Trails A Accuracy	OCS Trail A Accuracy	r(268)=.31, p< .001	MoCA Naming Accuracy	r(74)=.17, p=.154
Trails A Accuracy	CLQT A Accuracy	r(73)=.27, p=.018		
Trails A Time	OCS Trail A Time	r(110)=.38, p< .001	MoCA Naming Accuracy	r(74)=0.07, p=.532
Trails A Time	CLQT A Time	r(68)=.5, p< .001		
Trails B Accuracy	OCS Mixed Accuracy	r(265)=.4, p< .001	MoCA Naming Accuracy	r(74)=.06, p=.597
Trails B Accuracy	CLQT B Accuracy	r(73)=.49, p< .001		
Trails B Time	OCS Mixed Time	r(98)=.33, p< .001	MoCA Naming Accuracy	r(74)=0.01, p=.906
Trails B Time	CLQT B Time	r(67)=.5, p< .001		
Trails Exec. Score	OCS Executive Score	r(309)=0.17, p=.002	MoCA Naming Accuracy	r(74)=0.07, p=.558
Delayed Recall + Recognition Accuracy	MoCA Word Recall Accuracy	r(77)=.14, p=.22	RBANS Figure Copy Accuracy	r(71)=.05, p=.685
Delayed Recall + Recognition Accuracy	OCS Sentence Recall Accuracy	r(335)=.43, p< .001		

Table 2.3: Convergent correlations between OCS-Plus tests and construct- and format-matched neuropsychological validation tests *(continued)*

OCS-Plus task	Convergent test	Convergent Correlation	Divergent test	Divergent Correlation
Episodic Recognition Accuracy	OCS Episodic Memory Accuracy	r(338)=.24, p< .001	RBANS Figure Copy Accuracy	r(71)=.27, p=.023
Figure Copy Accuracy	RBANS Figure Copy Accuracy	r(69)=.66, p< .001	MoCA Naming Accuracy	r(70)=.33, p=.004
Figure Recall Accuracy	RBANS Figure Recall Accuracy	r(68)=.64, p< .001	MoCA Naming Accuracy	r(69)=.29, p=.014
Rule Finding Accuracy	Brixton Spatial Anticipation Test Accuracy	r(63)=.62, p< .001	MoCA Naming Accuracy	r(73)=.23, p=.045
Rule Finding Rules Learned	Brixton Spatial Anticipation Test Rules Learned	r(63)=.61, p< .001	MoCA Naming Accuracy	r(73)=.23, p=.051
Rule Finding Time	Brixton Spatial Anticipation Test Time	r(52)=.21, p=.134	MoCA Naming Accuracy	r(71)=.13, p=.256
Cancellation Accuracy	BIT Star Cancellation Accuracy	r(71)=.78, p< .001	MoCA Naming Accuracy	r(71)=.22, p=.061
Cancellation False Positives	BIT Star Cancellation False Positives	r(71)=.65, p< .001	MoCA Naming Accuracy	r(71)=.04, p=.707
Invisible Cancellation Accuracy	BIT Star Cancellation Accuracy	r(71)=.58, p< .001	MoCA Naming Accuracy	r(71)=.14, p=.224
Invisible Cancellation Correct Revisits	Corsi Block Accuracy	r(40)=0.1, p=.517	MoCA Naming Accuracy	r(71)=0.1, p=.4
Total Tasks Imapired	MoCA Total	r(78)=0.67, p< .001		

Incidence and Sensitivity Analyses

Next, I investigated OCS-Plus task impairment incidences. Figure 2.3 shows the task-by-task impairment incidence for each stroke cohort. The OCS-Plus invisible cancellation task showed the highest impairment incidence for both samples (subacute 85.16% impaired; chronic 76.73% impaired), whereas the semantics task showed the lowest impairment incidence (subacute 10.56% impaired; chronic 5.45% impaired). A numerical decrease in impairment rate was observed in the chronic sample, versus the subacute sample, for all measures. However, the Welch Two Sample t-test found no significant difference in incidence between the samples ($t(28) = 1.39,\ p = .18,\ d = -.51$).

Next, we investigated OCS-Plus sensitivity and specificity for each subtask on matched neuropsychological tasks. Table 2.4 presents results of these analyses. Simpler tasks (e.g., picture naming, semantics, orientation, word encoding, delayed recall and delayed recall and recognition, and episodic recognition) on the OCS-Plus had high specificity (>.80) at a cost of sensitivity (<.50), however, more complex tasks in the OCS-Plus (e.g., rule finding, figure copy/recall, and cancellation) which were designed to detect subtle deficits had exceptionally high sensitivity (>.90) and moderate to high specificity. Only the Invisible Cancellation task had a very low specificity (<.20), though had perfect sensitivity.

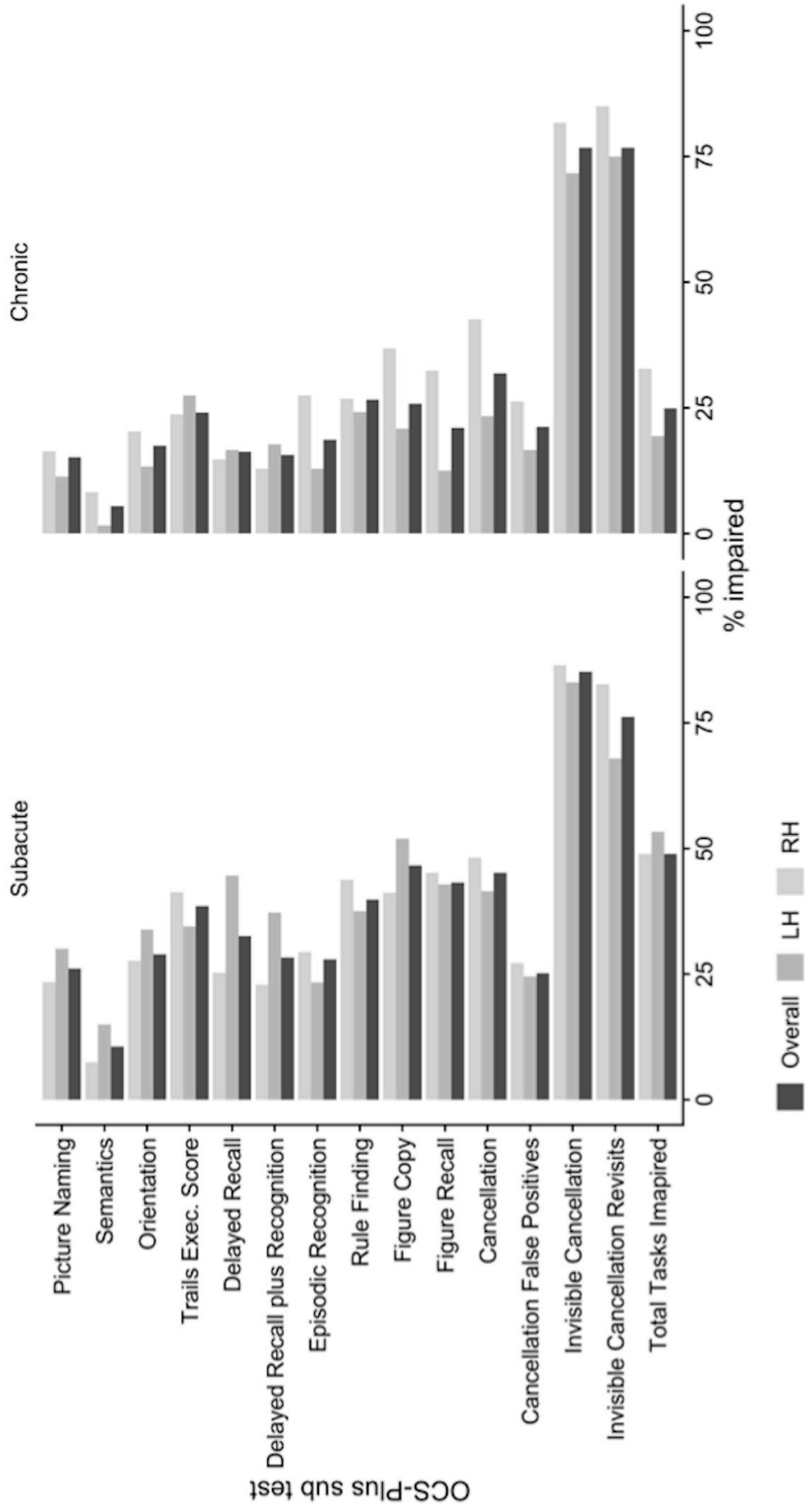


Figure 2.3. The proportion of stroke survivors classified as impaired on eachOCS-Plus subtask are shown separately for sub-acute(<3 months post stroke) and chronic stroke (>6 months post-stroke) survivors and left-hemisphere and right-hemisphere lesions. Figure available under CC-by 4.0 license https://osf.io/m3kc4/

Table 2.4: Comparison of the OCS-Plus sub-test impairment classifications to validation test impairment classifications in sub-acute (<3 months post-stroke) and chronic (>6 months post-stroke) stroke survivors

OCS-Plus Task	Validation Test	N	True positive	True negative	False positive	False negative	Sensitivity	Specificity
Picture Naming	OCS Picture Naming	342	11.40	64.04	9.65	14.91	43.33	86.90
Semantics	OCS Semantics	342	2.05	83.63	5.85	8.48	19.44	93.46
Orientation	OCS Orientation	337	11.28	63.50	12.46	12.76	46.91	83.59
Trails Exec. Score	CLQT Score	75	21.33	34.67	20.00	24.00	47.06	63.41
Delayed Recall	OCS Delayed Recall	295	14.92	53.22	8.81	23.05	39.29	85.79
Delayed Recall plus Recognition	OCS Delayed Recall plus Recognition	337	13.65	62.02	8.61	15.73	46.46	87.82
Episodic Recognition	OCS Episodic Recognition	340	8.24	63.82	15.59	12.35	40.00	80.37
Rule Finding	Brixton Errors	65	13.85	61.54	23.08	1.54	90.00	72.73
Figure Copy	RBANS Figure Copy	71	18.31	53.52	26.76	1.41	92.86	66.67
Figure Recall	RBANS Figure Recall	71	12.68	61.97	23.94	1.41	90.00	72.13
Cancellation	BIT Cancellation	73	28.77	58.90	9.59	2.74	91.30	86.00
Invisible Cancellation	Corsi Block	42	28.57	11.90	59.52	0.00	100.00	16.67
Proportion Tasks Impaired	MoCA impairment <26	80	85.00	0.00	12.50	2.50	97.14	0.00
Proportion Tasks Impaired	MoCA impairment <23	80	71.25	1.25	26.25	1.25	98.28	4.55

Next, I investigated OCS-Plus impairment classifications versus impairment classifications on clinically used first line cognitive screening tools (OCS and MoCA). Impairment classifications were determined based on previously published cut-offs detailed in methods. Overall, 87.50% of the sub-acute sample scored below MoCA cut-off of 26-points, and 72.50% below the 23-point cutoff. Of those classified as 'unimpaired' on MoCA, 100% were impaired in at least one OCS-Plus test. In terms of accuracy (calculated as $truepositives + truenegatives/(truepositives + truenegatives + falsepositives + facenegatives)$), compared to the MoCA 26 cut off, the OCS-Plus has an accuracy of 87.5%, and compared to the MoCA 23 cut off, an accuracy of 72.5%.

On the OCS, 96.22% of stroke survivors showed at least one cognitive impairment. Of the remaining 13 participants without an impairment on any of the OCS domains, 12 were impaired in at least one OCS-Plus test. This gives OCS-Plus a sensitivity of 98.5% vs OCS. Specificity could not be calculated for the MoCA vs OCS-Plus, as for all of the participants who did not have any tasks impaired on the OCS-Plus, they did not take part in the MoCA for comparison due to other non-project related reasons. Further analysis of sensitivity/specificity could not be conducted on the OCS as there is no value for impairment in number of tests currently.

2.1.6 Discussion

I generated new evidence for the psychometric of the OCS-Plus in a sub-acute and chronic stroke survivor cohort, and followed COSMIN criteria for reporting construct validity. First, I evidenced convergent construct validity for all OCS-Plus subtasks and divergent construct validity for all OCS-Plus subtasks, except the OCS-Plus language, figure cop, and orientation tasks, which related to visual moadlity assessments. Second, the OCS-Plus showed near perfect sensitivity for detecting subtle post-stroke cognitive impairments compared to two clinically used cognitive screening tools (MoCA and OCS) and acceptable test accuracy compared to the MoCA. Sepcificity could not be calculated in the current study for comparison to MoCA and OCS. Overall, I demonstrated that the OCS-Plus may be a valid and

sensitive cognitive screening tool for subtle post-stroke cognitive impairments, with sensitivity comparable to detailed neuropsychological assessments. By validating the OCS-Plus in a large real-world clinical rehabilitation sample and long-term chronic survivors, these results will be generalisable to a large extent. Limitations may apply in terms of comparability of demographic factors – such as age profiles of the stroke population (here, average age at stroke was 72 years) - and clinical factors – such as stroke severity (here, average NIHSS was 7.50). In an ideal world, comparisons would also be made longitudinally to compare impairment rates from acute to chronic time points.

For convergence, the unanticipated correlation between the OCS-Plus language subtasks with the visuo-spatial assessments may be explained by the visuo-spatial components of the OCS-Plus language subtasks. For example, the OCS-Plus Picture Naming task requires visual recognition of images. Therefore, stroke survivors with visuo-spatial deficits may struggle with this task, independent of their language ability. For the orientation subtask, the result may be spurious. To test the hypothesis that the visual modality of tasks was influencing the relationships, a simple correlation of OCS-Plus picture naming to the visual fields task in the OCS revealed a significant correlation ($r(340)$=.15, p = 003), albeit below convergence benchmarks.

Second, I showed that the OCS-Plus was more sensitive in detecting subtle cognitive impairments than other widely used cognitive screening tools (OCS and MoCA). Nevertheless, task-by-task sensitivity analyses indicated that tasks with a relatively small range of available scores – such as OCS-Plus Picture Naming, Semantics, and Orientation – had higher specificity than sensitivity. This suggests there is a trade-off in terms of task complexity and specificity/sensitivity. As such, more complex OCS-Plus tasks (e.g., Rule Finding) may be ideal for detecting subtle domain-general cognitive effects from broader vascular factors, linked to cognitive hallmarks of vascular dementia and small vessel disease (e.g., executive dysfunction [205]), whereas the simpler OCS-Plus tasks may be better suited for detecting core deficits (e.g., aphasia, orientation). By combining both types of tasks in OCS-plus,

the tool provides a time efficient approach to screening for both core cognitive impairments and more subtle vascularrelated post-stroke cognitive impairments. This provides a representative snapshot of post-stroke cognition, where initial domain-specific impairments may be improving or stable [206–208] and domain-general vascular and neurodegenerative factors may impede cognitive recovery [205].

It must be noted that there is a trade off for sensitivity in comparison to the OCS and MoCA, in that I found participants who were unimpaired on MoCA always had at least one impairment on OCS-Plus. It is possible that this is due to the differences in scoring; as the MoCA produces only an overall score, potentially doing worse than normative data on an individual subtask would not be shown in the overall score. For example, if an individual completely forgets all five words in the delayed recall, and show a clear recall problem, they would still pass the MoCA if all other tasks were perfect. In this way, impairments are being masked in the MoCA. On the other hand, perhaps the way the OCS-Plus is scored by subtask, means that one is more likely to find an *impairment* but really it is a result of having multiple tests for one individual. For example, some research has shown that the more tests you give someone the more likely they are to *fail* one, not because they are necessarily impaired but due to chance or fale-positives through numbers [209]. Further research must examine this further and perhaps compare subtask scoring to an overall score for OCS-Plus in comparison to neuropsychological batteries. However, by default, a neuropsychological battery may fall at the same hurdle as OCS-Plus and falsely classify an individual as impaired if they complete enough tasks.

Third, our data found numerically, but not statistically, lower incidence of cognitive impairments in the chronic stages post-stroke compared to the earlier subacute stage. Nevertheless, impairment prevalence remained high for the invisible cancellation task, a sensitive test of working memory. It may be that these chronic working memory deficits reflect not only stroke-related damage, but also more subtle vascular-related damage that accrues during ageing. As such, these deficits may be less amenable to recovery after stroke, which may explain the consistently high impairment prevalence in the chronic stage post-stroke. The higher rate

of chronic impairment for the Invisible Cancellation subtask, rather than other memory subtasks, may be explained by its increased load on working memory. More specifically, the Invisible Cancellation subtask requires participants to store in working memory which targets have already been selected and their location, which loads more heavily onto working memory than simply recalling a figure or words.

The OCS-Plus offers several advantages over currently used paper-based clinical screening tools and neuropsychological test batteries. First, the OCS-Plus report gives clear information about both domain-general and domain-specific cognitive performance, in contrast to traditional paper-based screening tools such as the MMSE and MoCA, which provide a coarser evaluation of cognitive functioning overall [37, 48]. Secondly, the OCS-Plus is available on a platform independent app that provides standardised administration instructions for the user and automatically scores participants against age-adjusted impairment cut-offs. This contrasts with currently used tools that require manual test scoring [37, 48]. Manual scoring may increase the time burden of administration and may also increase error in the scoring process, relative to automated approaches. Though it must be noted, a paid for app version of the MoCA now exists. Thirdly, the OCS-Plus takes on average 24 minutes to administer [50] and thus offers substantial time advantages relative to extensive neuropsychological test batteries, which can take upwards of one hour to administer.

These features of the OCS-Plus should be considered in the context of clinical practice. Firstly, clinicians could use fine-grained information about domain-specific and domain-general cognitive functioning – in addition to other factors – to detail prognosis and recovery, and aid conversations around adjustment to living life post-stroke [210]. Secondly, as the OCS-Plus app provides standardised administration instructions, it could be used in clinical practice by a range of allied health professionals, without specific neuropsychology training, which is required to administer neuropsychological test batteries. Thirdly, where time is pressured, having an automatically scored tool could accelerate the assessment process and return crucial time to clinicians for other aspects of assessment, providing potential cost savings. Finally, a tool that is quick to administer and

highlights subtle cognitive deficits [50] may be a valuable adjunct in discharge planning and education of patients. Overall, these factors suggest that the OCS-Plus may be a valuable cognitive screening tool for use in clinical populations, particularly in stroke survivors who may present with a mixture of domain-specific and domain-general cognitive changes.

Several limitations should be noted with regards to both the present study method and the OCS-Plus tool itself. With regards to the study method, the analyses contain different sample sizes as some participants did not complete all sessions, due to both patient-specific and environmental factors. This made it difficult to correct for multiple comparisons, and I regret that statistical power varies between analyses. Second, most participants completed tests across several brief sessions, due to factors such as fatigue. Post-stroke cognition is not constant but dynamically changing, and stroke survivors' cognitive abilities may have fluctuated over these intervals. I attempted to mitigate this issue by ensuring that all validation tasks were completed within a maximum period of 30 days. With regards to limitations of the OCS-Plus tool itself: in some tasks, the images may have insufficient contrast for patients with pre-existing visual impairments, such as macular degeneration (e.g. see Rowe & Hepworth [211]), and this may be exacerbated on a reflective tablet surface. Practical issues associated with tablet-based testing like running out of charge, or cracks on the screen may further impact testing. Nevertheless, these issues can be considered relatively minor. Therefore, they should not – in theory - impede the use of the OCS-Plus in research studies and/or clinical practice. All cognitive screening tools are imperfect, a further consideration aside from the strengths and weaknesses reported here could be patient preference. No research has thus far compared different cognitive tasks and analysed feedback on each, this is worthy of future investigation to further evidence choice of tool for clinicians.

Overall, the OCS-Plus may be a valid and sensitive cognitive screening tool which includes detecting more subtle cognitive impairment in stroke survivors. Indeed, the OCS-Plus was found to detect cognitive impairments in a large sample of subacute and chronic stroke survivors at a similar level to selected standardized and validated

neuropsychological tests, while offering substantial practical and time advantages over these tests. As such, the OCS-Plus could be considered for implementation in clinical practice. There are significant limitations regarding specific subtasks and their ability to detect impairment, but we have detailed inferences that can be taken in terms of sensitivity and specificity in light of these. Future research could attempt to disentangle domain-specific and domain-general cognition trajectories and underlying neuroanatomical correlates using OCS-Plus. In addition, the validity of using the OCS-Plus in different clinical cohorts, which may similarly require more sensitive cognitive screening, should be investigated. Furthermore, comparing performance on OCS-Plus to the perspective of impairment by stroke survivors would aid in evidencing the clinical meaningfullness of OCS-Plus outcomes.

3

Ecological and predictive validity of Multiple Errands Tests

3.1 Comparing the Oxford Digital Multiple Errands Test (OxMET) to a real-life version: convergence, feasibility, and acceptability.

This section is re-written from an accepted for publication manuscript: **Webb, S. S** *., & Demeyere, N. (2024). Comparing the Oxford Digital Multiple Errands Test (OxMET) to a real-life version: convergence, feasibility, and acceptability. Neuropsychological Rehabilitation*

3.1.1 Abstract

Aims I aimed to assess the convergence, feasibility, and acceptability of the Oxford Digital Multiple Errands Test (OxMET) and the in-person Multiple Errands Test – Home version (MET-Home). **Methods** Participants completed OxMET, MET-Home, Montreal Cognitive Assessment (MoCA), and questionnaires on activities of daily living, depression, technology usage, mobility, and disability. **Results** 48 stroke survivors (mean age 69.61, 41.67% female, and avg. 16.5 months posts-stroke) and 50 controls (mean age 71.46, 56.00% female) took part. No performance differences

were found for healthy and stroke participants for MET-Home, and only found below p = .05 for OxMET but not below the corrected p = .006. Convergent validity was found between MET-Home and OxMET metrics (most r>=.30 & p<.006). MET-Home accuracy was related to age (B= -0.04, p= .03), sex (B= -.98, p= .03), disability (B= -0.63, p= .04), and MoCA (B= .26, p<.001), whereas OxMET accuracy was predicted by MoCA score (B=.40, p<.001). Feedback indicated that the OxMET was easy and fun and more acceptable than the MET-Home. The MET-Home was more stressful yet interesting. **Conclusions** The MET tasks demonstrated good convergent validity, with the OxMET digital administration providing a more feasible, inclusive, and acceptable assessment, especially to people with mobility restrictions and more severe stroke.

3.1.2 Introduction

Executive dysfunction is defined as a deficit in higher level cognitive abilities such as planning and self-monitoring behavior, switching tasks or attention between tasks, and updating current schema to adjust behavior [109, 212]. Executive dysfunction has been found to be a core domain relating to functional outcomes [15, 213, 214]. Accurate identification of executive dysfunction could reduce the burden on health services through early intervention [215]. Clinical decisions regarding, for example, patient discharge and care packages should therefore take into account cognitive functioning and the impact of any cognitive impairments on ADLs [216].

The Oxford Digital Multiple Errands Test (OxMET) is a MET adaptation that has been developed for use on a computer tablet [130], making it standardized for all participants and more inclusive for participants with mobility problems. Mobility is an important factor to consider in the context of stroke rehabilitation, especially as movement adds an additional cognitive load to daily activities [13]. The OxMET does not seek to remove the need to assess the added load of motor impairments, but instead the present investigation, and overall OxMET purpose, is to assess performance on a real-world complex executive multiple errands task, including in those with severe mobility restrictions. Furthermore, the OxMET can

be completed in less than five minutes, and thus has cost and time advantages compared to real-life MET versions.

The OxMET has been validated in stroke, and was demonstrated to be a reliable measure of executive function with good sensitivity and known-group discrimination, and moderate convergence with executive functioning tests [130]. Convergence was above the convergence benchmark of r =.30 [129]. Up to now, OxMET had not been compared to other MET versions to establish its relationship to observed in-person activities of daily living. This comparison is critical to determine whether the digital OxMET shows convergent validity to the traditional in-person format of a multiple errands task, but also to examine whether it indeed offers advantages in terms of its feasibility and acceptability.

The aim of this first project was to compare the OxMET to the MET-Home, a previously validated home-based in-person version of the MET, to examine the construct validity, acceptability, and feasibility of the OxMET and MET-Home. The MET-Home was chosen as it was the most adaptable English-language in-person MET to administer in home visits including in rural Oxfordshire. In particular, the MET-Home does not require moving participants from their residence (i.e., whether homes or nursing homes etc). As part of this project, I developed and tested a revision of the MET-Home which is more acceptable to stroke survivors. I document this revision process in the upcoming publication in the appendix. I used a mixed-methods approach to examine:
1. Associations between the OxMET and the MET-Home. 2. Completion rates and feasibility of the OxMET and MET-Home. 3. Participant feedback on acceptability.

3.1.3 Methods

How the sample size was determined, any data exclusions (if any), all manipulations, and all measures in the study are reported (Solution for transparency by Simmons et al. [171]). The COSMIN guideline for studies on measurement properties [217] and the STROBE cohort checklist [188] were adhered to in this manuscript. Ethical approval was granted by the Medical Sciences Interdivisional Research

Ethics Committee (REF: R73921/RE001) of the University of Oxford. Written informed consent was provided by all participants.

Participant recruitment

The present study recruited chronic stroke survivors (approximately 6 months post-stroke) from the research volunteer database including people who had previously taken part in the Oxford Cognitive Screening studies (OCS-RECOVERY [IRAS project ID: 241571] and OX-CHRONIC [IRAS project ID: 259478]). All stroke survivors were based in the community at time of recruitment.

Neurologically healthy controls were recruited either from the healthy ageing research volunteer database or were family/friends/partners of the stroke survivors in the study. Healthy controls were included here for increased statistical power (see Webb & Demeyere, [218] for a discussion on samples in neuropsychological research).

Sampling for both samples was based on convenience (those who replied to the contact, those who were partners/family of existing participants) and snowballing sampling (word of mouth from existing participants) methods.

Stroke survivors who were recruited met the following inclusion criteria: 1) confirmed clinical stroke diagnosis from medical notes, 2) able to concentrate for at least 20 minutes (judged by the participant), and 3) able to give informed consent (mental capacity assessed as part of consent process following approved protocol). Neurologically healthy controls met the following inclusion criteria: 1) no self-reported neurological or psychiatric complaints or diagnoses; and 2) a Montreal Cognitive Assessment (MoCA) score >22 [37] on the day of participation.

Exclusion criteria for both groups included sensory/perceptual/motor impairments that would prevent the ability to complete the tasks beyond reasonable adjustment (not inclusive of wheelchair/assistance use, which still allowed participants to complete the tasks). A priori power calculations indicated a minimum of 66 participants for the convergent validity correlation analyses (alpha = .05, power = 80%, one sided, correlation >.30) and a minimum sample size of 99 participants

for the planned linear regression analysis (corrected alpha = .05, power = 80%, f^2 = .15[medium effect], and eight covariates).

Stroke survivors were included without regard to cognitive ability or executive functioning difficulties to better reflect the clinical reality of any given presenting stroke survivor. I did not aim to assess the executive dysfunction detection abilities of either MET in the current study.

Measures

MET-Home The MET-Home is a standardized and validated assessment version of the Multiple Errands Test [106] which requires participants to complete fourteen household tasks in their place of residence [126]. Examples of tasks include watering a plant and noting the cost with delivery of a two-item pizza. Participants are instructed to complete all tasks in any order as fast as possible but without rushing. They must follow several rules while completing the tasks: 1) participants must stay on the property; 2) participants must not go back to a room they have already visited, except where it is the only route, 3) participants must not speak to the test administrator, except to ask the time for one of the tasks, and 4) participants must not collect items, but rather identify them. Further detail about the MET-Home, including in depth scoring methodology, is reported elsewhere [126], at www.met-home.com, and in Table 7.8 in the appendix. In brief, the MET-home is scored based on summed number of tasks completed (accuracy), summed number of partially completed tasks, and summed number of omitted tasks from the errands list, as well as additional error metrics. The task administrator read the scripted instructions for the MET-Home and asked participants to complete the task.

OxMET The OxMET is a simplified, standardized, normed, and validated computer-tablet version of the Multiple Errands Test[106] that requires participants to buy six items and answer two questions by going through a street of shops on a screen – it is described in detail elsewhere [130]. There are rules to follow including to take as little time as possible, to spend as little money as possible,

to only enter a shop to buy something or to answer a question, and to only visit a shop once. The scoring of the OxMET is automatic and requires no examiner input. Scoring primarily involves one outcome metric of task accuracy, which is the sum of tasks accurately completed minus errors made (score range -10 to 10), as well as additional error metrics detailed in the original investigation [130] and in Table 7.8 in the appendix.

Attempts and completions Participants were classified into one of the following categories (these novel categories were determined a priori by the research team to provide quantitative data on acceptability and feasibility of the two tests): 1) task attempted and completed; 2) task attempted but not completed, or 3) task not attempted. 'Attempted' was defined as a willingness to attempt the task. 'Completion' was determined by the researcher, and where appropriate, self-reported by the participant. That is, for the task to be 'completed', either participants reported that they had finished, or it was determined by the examiner based on observation. Task completion was independent of accuracy on any item or the last item regarding self-reported completion for the MET-Home[1].

A similar categorization of attempts and completions for the OxMET was applied: attempting was counted as reading through the instructions, non-completion was where participants wanted to stop after having started the task. Completion was defined as either self-reported completion of the task, again irrespective of accuracy/errors, or the researcher would prompt the participant to ask if they were finished after a reasonable time where it became obvious that the individual was finished.

[1]If participants did not explicitly report being finished with the task (i.e. they failed the last item of the MET-Home requiring them to self-report being finished), the researcher would prompt the participant to ask if they were finished after a reasonable time where they observed that the participant appeared to have finished with the task. For example, for the MET-Home, the participant had sat back down, put the papers aside, and appeared to be awaiting more instructions. Or for the OxMET, participants placed the tablet down and looked at the researcher.

Non-MET measures I collected clinical, demographic, and functional outcome measures, including disability/independence related data. Demographics included age at testing, years of formal education (i.e., each year of full-time formal education [with a teacher] was counted as a year, with part time teaching being divided by half to determine full time equivalent), dominant hand (self-reported and prior to stroke if relevant), sex, as well as stroke information obtained from clinical notes.

For all participants I collected demographics and screened for global cognitive functioning with the Montreal Cognitive Assessment (MoCA) [37]. The MoCA, is one of the most commonly used cognitive screening tool in stroke [45] and ranged 0-30 with higher scores reflecting better performance. I used the Geriatric Depression Scale (short form - 15 item) (GDS-15) [219] to measure depression severity to account for the influence of mood on cognitive assessment and functional abilities [220, 221], where higher scores reflect greater severity of depression. Basic Activities of Daily Life (ADL) performance measured with the most commonly used [222] functional independence measure, the Barthel Index [223] (range of 0-20),), and for instrumental ADLs I used the most commonly used measure in stroke trials [224], the Nottingham Extended Activities of Daily Living scale (NEADL) [225] (range of 0-66). For both ADL measures, higher scores reflect more independence. I additionally classified stroke severity at stroke admission using the National Institute of Health Stroke Scale (NIHSS) [226] and disability/dependence using the Modified Rankin Scale (mRS) [227] which is the most commonly used outcome measure in stroke trials [228].

For the majority of participants (n=80), I collected information on the frequency of technology use using the Computer Gesture Therapy for Aphasia battery Technology Screen questionnaire, which is intended as a novel, validated, and aphasia friendly, assessment suitable for stroke survivors [229]. The scale on confidence in using technology is included in the technology screen that was adapted from an existing scale [230]. The confidence scale was included later into the project to assess the degree to which technology usage may predict completion and accuracy on the OxMET as a computer tablet-based task.

To minimize the sample bias in the project I did not pre-select participants based on any criteria other than willingness to take part and eligibility. This means I selected stroke survivors who fit the inclusion criteria, and I did not use further criteria such as stroke severity or cognitive ability. Some level of research recruitment bias is unavoidable, in that those who are willing to take part in the project may be different to those who did not sign up, and that I only recruited those who could consent to take part.

Procedure

During the in-person research visit at participant homes, following informed consent, I administered the mRS, Barthel Index, NEADL, GDS-15, and the technology screen. Following this, I administered the MoCA. Finally, I administered the OxMET and MET-Home in counterbalanced order.

The duration of the research visit and full protocol was typically 1hr-1.5hrs. I conducted all data collection, with additional help from two research assistants for five participants (total N=98). Training for MET-Home was completed via training materials on the MET-Home website and via communication with the MET-Home author. I was trained and licensed to administer and interpret the MoCA. After completing each MET, I asked an open question to probe feedback on the task: 'Do you have any feedback about the tasks, anything you liked or didn't like or anything else you would like to remark upon?'. I recorded feedback on case record forms for each participant at the end of the session. Audio recordings were not made. I did not conduct a full qualitative analysis for participant feedback but summarize the feedback descriptively.

Data analysis

All data analysis was conducted over both sample groups. First, I present descriptive statistics and inferential statistics on sample differences in performance on each MET. Next, I correlated all OxMET to MET-Home metrics. Convergence overlap was determined as correlations above r=.30 where this benchmark is used in

MET-literature [129]. I examined the feasibility of administering both MET tasks: For all participants, I included age, education, sex, mRS, mobility, tech-screen usage, MoCA score, and GDS-15 score as predictors of OxMET completion (0 or 1 for completion, 1 being completion)/accuracy using binomial logistic and Gaussian linear regression and ran the same model without the tech-screen for MET-Home completion/accuracy. I added time since stroke and stroke severity in separate identical models for stroke survivors. Missing data were not imputed, all analyses were conducted on complete data only. Missing data was sufficiently low that using only complete cases for each analysis provided enough statistical power. We briefly examined health economic considerations for using either task in clinical practice. I calculated staff costs per MET from the Schedule of Events Cost Attribution Template A calculation from the National Institute for Health and Care Research. I input the average completion time rounded to zero decimal places and use of medical staff for one person for one visit. All analyses scripts and data are available in the project repository (`https://osf.io/pzwgf`).

Brief analysis of feedback Data were analysed between NVivo 12 for Mac and Nvivo for Windows. Instead of participant names I used pseudo-anonymized participant codes divided into either healthy controls or stroke survivors. I used a qualitative description approach to understanding and presenting the feedback data [231]. The approach aims to describe the experience under examination in a way that is easily understood and based primarily on direct quotes or statements [231]. I coded responses using a basic thematic analysis approach but because I did not record responses verbatim in the first part of the study, I only report main, high frequency themes/phrases/statements of meaning.

To explore whether feedback valance was in any way related to performance on the MET tasks (e.g., those who found that tasks easy, thought more positively of the tasks), I regressed accuracy to predict binarized positive or negatively valanced feedback.

3.1.4 Results

Participants

A total of 150 participants were contacted, of these, a total of 98 participants were convenience sampled. Reasons for non-participation are presented in supplementary materials and further detailed in the later feasibility section. The final sample included 48 stroke survivors who were recruited from the community and 50 healthy controls. I present the summary demographics of both healthy control and combined stroke survivor samples, as well as percentage of missing data per N, in Table 3.1. I ran Welch two sample t-tests to compare age and education demographics between samples for use as covariates in later analyses; I found a significant difference in education ($t(93.54)$, 2.43, p=.01, 95% CI [.32, 3.16]) but not age ($t(88.86)$, .72, p=.472, 95% CI [-3.25, 6.96]). I account for education differences in later group comparison analyses.

Table 3.1: Summary of sample characteristics for all participants (healthy controls n =50, stroke n = 48)

	Healthy Controls		Stroke	
Characteristic	N	Value	N	Value
Age (M(SD))	50 (0%)	71.46 (11.03)	48 (0%)	69.61 (14.14)
Education (M(SD))	50 (0%)	15.68 (3.32)	48 (0%)	13.94 (3.75)
Handedness	50 (0%)	A= 4%; L= 8%; R= 88%;	48 (0%)	A= 2.08%; L= 14.58%; R= 83.33%;
Sex	50 (0%)	M= 44%; F= 56%;	48 (0%)	F= 41.67%; M= 58.33%;
Ethnicity	50 (0%)	Asian or Asian British: Indian= 2%; White: Any other white background= 4%; White: English, Welsh, Scottish, Northern Irish or British= 94%;	48 (0%)	Black, Black British, Caribbean, or African: Any other Black, Black British, Caribbean, or African background= 2.08%; Other ethnic group: Any other ethnic group= 2.08%; White: Any other white background= 8.33%; White: English, Welsh, Scottish, Northern Irish or British= 87.5%;
Days Since Stroke (M(SD, range))	-	-	48 (0%)	502.29 (547.47, 122-3283)
Stroke type	-	-	48 (0%)	subarachnoid haemorrhage= 2.08%; TIA= 6.25%; intracerebral haemorrhage= 12.5%; ischaemic= 79.17%;
Stroke side	-	-	48 (0%)	Undetermined= 6.25%; B= 12.5%; L= 33.33%; R= 47.92%;
Stroke Severity (Median(IQR))	-	-	43 (10%)	6 (4-9.5)

Table 3.1: Summary of sample characteristics for all participants (healthy controls n =50, stroke n = 48) *(continued)*

Characteristic	N	Value	N	Value
Mobility	50 (0%)	Walks independently but with difficulty (may use aid, e.g., frame/stick)= 4%; Walks completely independently with no difficulty= 96%;	47 (2%)	No sitting balance= 2.13%; Wheelchair + sitting balance= 2.13%; Walks with assistance of 1= 10.64%; Walks completely independently with no difficulty= 40.43%; Walks independently but with difficulty (may use aid, e.g., frame/stick)= 44.68%;
Modified Rankin Scale (M(SD))	50 (0%)	1= 2%; 2= 2%; 0= 96%;	48 (0%)	5= 2.08%; 0= 8.33%; 4= 10.42%; 1= 18.75%; 2= 27.08%; 3= 33.33%;

Descriptive statistics

In Table 3.2 I present descriptive statistics for each metric from both the MET-Home and OxMET and non-MET measures by sample group, with statistical ANCOVA comparisons controlling for differences in education. Education and age are both known to influence performance [126, 130]. There were no statistical differences found for healthy controls versus stroke survivors on the MET-Home. For the OxMET, the sample groups were statistically different on accuracy, frequency of rule breaks, omissions, time taken to complete the task, and total errors. No differences in groups on either MET survived corrections for multiple comparisons, though effect sizes were significantly higher for OxMET sample differences (Welch Two Sample t-test on effect sizes found: $t(11.38)=2.44$, $p=.03$). Non-MET measures were all statistically significant below .001 and further justifies use as covariates in analyses.

Convergent Associations between OxMET and MET-Home

I correlated all metrics from both METs using Spearman's Rho analysis due to data type not being continuous and not normally distributed. The results of the convergent analyses are presented in Table 3.3 with correction for multiple comparisons ($**=p<.00625$, $*=p<.05$). All analyses are presented for total transparency, though I focus on like-for-like association (e.g., accuracy on OxMET and accuracy on MET-Home)

Table 3.2: Descriptive statistics and ANCOVA comparisons between sample groups on MET-Home and OxMET as well as additional covariates

Measure	Metrics	Group	M	SD	Df1	Df2	F	p	ef
MET-Home	Accuracy	Control	10.10	2.27	1	88	2.65	.11	0.03
		Stroke	9.05	3.04					
	Freq. Passes	Control	2.74	2.13	1	88	0.30	.59	0.00
		Stroke	2.90	2.08					
	Freq. Rule Breaks	Control	8.10	4.32	1	88	0.89	.35	0.01
		Stroke	9.07	5.02					
	Inefficiencies	Control	0.26	0.60	1	88	0.54	.46	0.01
		Stroke	0.32	0.82					
	Number Rules Broken	Control	2.30	0.71	1	88	0.34	.56	0.00
		Stroke	2.39	0.97					
	Omissions	Control	1.40	1.69	1	88	3.15	.08	0.04
		Stroke	2.37	2.47					
	Partial Completions	Control	2.48	1.82	1	88	0.11	.74	0.00
		Stroke	2.54	1.57					
	Rooms Visited	Control	7.08	3.44	1	82	2.68	.11	0.03
		Stroke	5.68	3.01					
OxMET	Accuracy	Control	8.14	2.25	1	92	4.83	.031*	0.05
		Stroke	6.42	3.66					
	Commissions	Control	0.66	0.96	1	92	0.46	.497	0.00
		Stroke	0.91	1.38					
	Freq. Rule Breaks	Control	1.72	2.12	1	92	4.72	.03*	0.05
		Stroke	3.33	3.61					
	Omissions	Control	1.02	1.45	1	92	7.33	.01*	0.07
		Stroke	2.42	2.62					
	Partial Omissions	Control	0.52	0.79	1	92	1.97	.164	0.02
		Stroke	0.84	1.21					
	Perseveration	Control	0.04	0.20	1	92	1.06	.305	0.01
		Stroke	0.13	0.50					
	Time	Control	174.09	53.87	1	92	5.64	.02*	0.06
		Stroke	209.68	79.27					
	Total Error	Control	2.22	2.71	1	92	4.56	.04*	0.05
		Stroke	4.18	4.49					
Non-MET	Barthel	Control	19.78	0.62	1	95	9.70	<.001**	0.09
		Stroke	17.83	3.79					
	GDS15	Control	1.31	1.73	1	92	21.86	<.001**	0.19
		Stroke	4.48	3.99					
	MoCA	Control	26.24	2.38	1	93	17.59	<.001**	0.16
		Stroke	22.46	4.89					
	NEADL Total	Control	63.32	3.53	1	95	48.66	<.001**	0.34
		Stroke	44.23	17.60					
	Tech Screen Total Used	Control	11.79	2.57	1	77	8.51	0.01**	0.10
		Stroke	9.32	3.81					
	mRS	Control	0.06	0.31	1	95	141.52	<.001**	0.60
		Stroke	2.25	1.19					

[a] Note. ANCOVA included comparing group on score metric whilst accounting for variance explained by education.. "**" refers to significance on or below .006 (Bonferroni corrected per OxMET and MET-Home measure), and "*" refers to significance below .05. 'ef' refers to generalized effect size estimate.

Table 3.3: Spearman's Rho correlations between OxMET and MET-Home performance metrics inclusive of healthy controls and stroke survivors.

MET-Home Metrics	OxMET metrics							
	Time	Accuracy	Omissions	Partial Omissions	Commissions	Perseveration	Freq. Rule Breaks	Total Errors
Planning time	0.41**	-0.10	0.12	-0.02	0.10	-0.01	0.11	0.06
Time	0.25*	0.23*	-0.25*	-0.30**	-0.20	0.05	-0.25*	-0.28*
Accuracy	-0.19	0.37**	-0.37**	-0.26*	-0.26*	-0.01	-0.36**	-0.36**
Partial Omissions	0.07	-0.07	0.10	0.06	0.05	-0.08	0.08	0.08
Omissions	0.12	-0.38**	0.35**	0.24*	0.24*	-0.02	0.35**	0.35**
Num. Rule Breaks	0.03	0.08	-0.11	-0.24*	-0.13	-0.11	-0.13	-0.15
Freq. Rule Breaks	0.18	-0.07	0.04	-0.15	-0.01	0.06	0.04	0.00
Freq. Passes	0.24*	-0.12	0.11	0.09	0.10	-0.03	0.09	0.09
Inefficiencies	0.00	0.02	0.01	-0.12	-0.03	-0.11	-0.01	-0.07
Rooms Visited	-0.19	0.25*	-0.26*	-0.29*	-0.17	-0.22*	-0.26*	-0.29*

[a] Note. **=p<.00625, *=p<.05

Feasibility

Descriptives on completion of the OxMET versus MET-Home Reasons for non-participation in the study overall are presented in supplementary materials. Notably, one stroke survivor didn't want to be seen at home so did not want to take part at all and one stroke survivor did not have anywhere they felt they could complete the tasks due to their living situation. There were three stroke survivors who reported mobility was too poor to attempt the MET-Home, so they did not want to take part at all. Table 3.4 shows the attempt and completion rates of the OxMET and MET-Home in the same session. All healthy controls completed both tasks, and numbers deviating from completing both tasks refer to stroke survivors only.

For the one participant who agreed to take part in the study but did not attempt either the OxMET or MET-Home, they reported this was due to fatigue that they suddenly experienced after following the typical protocol. For the one participant who attempted and completed the MET-Home but did not do the OxMET, they were administered the MET-Home first and reported fatigue as reason for not attempting OxMET. No participants failed to complete the OxMET once attempted. For the four participants who attempted and completed the OxMET but did not attempt or complete the MET-Home, two participants refused due to low mobility, one refused due to fatigue (they were administered the OxMET first), and one refused as they felt their vision was too poor to read the text on the standard MET-Home instruction sheet. As the poor vision was a new development since recruiting the participant to the overall protocols, the researchers were unable to adjust the size of the document prior to travelling to their home.

For the three participants who attempted the MET-Home but did not complete it while having completed the OxMET: one participant realized they couldn't attempt the MET-Home as they lived in a shared home and did not want to go up and downstairs to their room, nor did they want the researcher to come with them whilst they moved in the shared home; the second participant who attempted but did not complete the MET-Home (MET-Home before OxMET) appeared confused as to what the requirements of the test were, and read the instructions and said

they finished after reading the instructions; finally, one participant started the task then after finishing the planning stage declined to continue due to low mobility. Due to the criteria for completion and attempts, they were classified as not completed.

Table 3.4: Summary of attempt and completion of the OxMET and MET-Home, assuming OxMET was attempted in all cases

	MET-Home		
OxMET	Not attempted	Attempted and not completed	Attempted and completed
OxMET not attempted	1	0	1
OxMET attempted and not completed	0	0	0
OxMET attempted and completed	4	3	89

[a] Note. Three participants completed the OxMET on a 6.5 inch smartphone and not a computer tablet due to equipment failure. They were included having completed the task on the smartphone, though the OxMET is recommended not to be used on a screen smaller than 7.5inch.

Predicting performance completion and accuracy The binomial logistic regression model of completion of the tests controlling for age, education, sex, mRS, mobility, tech-screen usage, MoCA score, and GDS-15 score could not be reliably interpreted due to low variance in completion rates (n=2 non-completed for OxMET and n=7 for MET-Home).

Using multiple linear regression with the same variables as the logistic regression but to predict task accuracy: for the OxMET, accuracy was predicted by MoCA score B=.40, p<.001 ($R^2_{adjusted}$=.38, $F(8,67)$= 6.74, p<.001). MET-Home accuracy was predicted by Age B= -0.04, p= .03, sex B= -.98 (males worse), p= .03, mRS B= -0.63, p= .04, and MoCA score B= .26, p<.001, but only the MoCA prediction survived correction for multiple comparisons ($R^2_{adjusted}$=.37, $F(7,81)$= 8.47, p<.001). Using stroke survivors only to include stroke severity and time since stroke: OxMET accuracy was predicted by MoCA score B=.47, p=.006 (at the alpha corrected significance level p<.006) but the overall model was not significant ($R^2_{adjusted}$=.29, $F(10,20)$= 2.25, p=.059) possibly due to low power, and MET-Home accuracy was predicted by NIHSS B= -0.20, p= .017 and MoCA score =.29, p=.005

($R^2_{adjusted}$=.50, $F(8,27)$= 5.50, p<.001), but none survived corrections for multiple comparisons. See Figure 3.1 for a visualization of predictors of task accuracy.

Brief health economics considerations

The OxMET was quicker to administer than the MET-Home, taking on average 3.19mins versus 11.21mins for MET-Home. The OxMET therefore would incur a lower cost in staff time (£8.82 versus £32.34). Scoring of the OxMET is automatically generated, incurring no staff time costs. Data on staff time required for scoring of the MET-Home, or writing up of any clinical reports for either MET was not collected as part of the study.

If a healthcare provider administering the OxMET does not readily have access to computer-tablet devices, there is an additional cost of approximately £366.55 for a computer-tablet [232], whereas the MET-Home is a freely available resource and only requires the printing of materials and purchasing or usage of at least one clip board, strap, and pen/pencil. For equivalence, the costs of acquiring a new tablet to run the OxMET, if one is not available to clinicians, is equivalent to administering the MET-Home to 12 participants ((£366.25 tablet + £8.82 staff time) / £32.34 staff time for MET-Home).

There are potential safety costs associated with in-person assessments that require physical activity, such as the MET-Home, particularly given the task requires participants to move around their home. The safety costs include participants who are at a falls risk, or who may not be, but do fall or otherwise injure themselves whilst maneuvering the home. In this project I observed one participant rushing during the task and almost falling down the stairs but thankfully caught themselves, which presented a clear hazard. Should the one participant have fallen, it may have cost a minimum of £77 if they were seen as an out-patient, or a single visit to an accident and emergency department would have cost £83, and going forward should they require community hospital assistance the minimum cost is £3,209 [233]. At least a third of older adults in the UK experience falls [234] and stroke survivors are

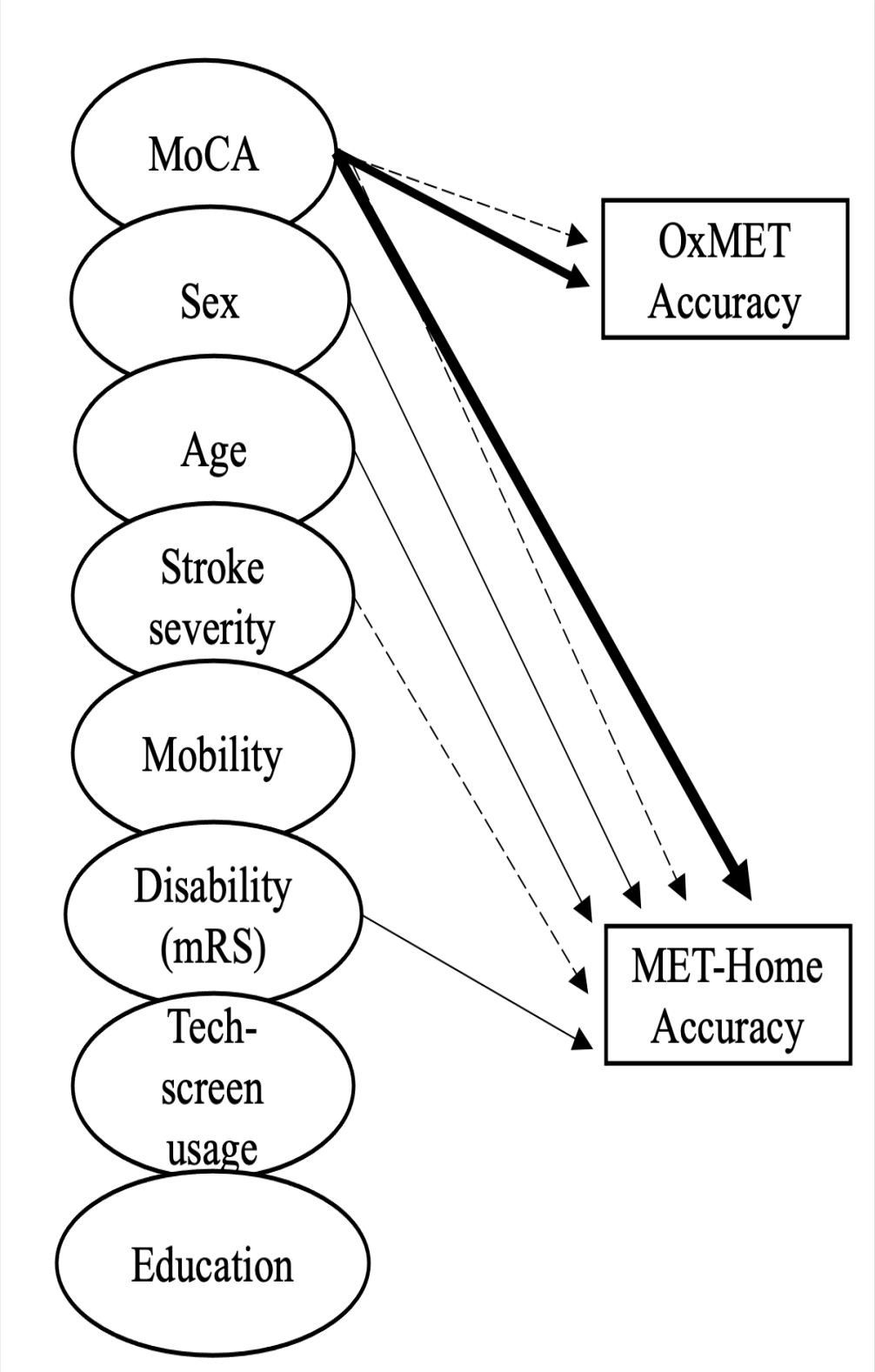


Figure 3.1: Visualization of statistically significant predictive associations across all participants (solid lines) and only stroke survivors (dashed lines) for task accuracy. Thicker lines are associations which survived corrections for multiple comparison ($p<.006$). Figure available under CC-BY 4.0 attribution license (https://osf.io/xw3s8)

at an even higher increased risk with 40 to 70% of stroke survivors experiencing a fall each year [235], and 14% experiencing one within the first month post-stroke [236].

Inferring acceptability from participant feedback

From the total sample, 85 participants provided feedback for the OxMET, and 86 gave feedback for the MET-Home. A full breakdown of feedback in the form of a code book exported from NVivo can be found in Table 7.5 in the supplementary materials. Feedback was not recorded for 8 participants due to examiner error. Following a summary of the feedback, I briefly examined whether accuracy on a test was associated with positive or negatively valanced comments on either test.

MET-Home feedback was more frequent (n=121) than the OxMET feedback (n=99) numerically. In terms of feedback valence, the majority of comments were positively worded (n=72 comments), followed by 54 neutral comments and 35 negative comments.

Participant feedback often regarded the METs as easy and fine (40 references). The vast majority of these references stated synonyms of 'easy'/ 'easy enough', 'straight forward', and 'fine.' Regarding the MET-Home, Control24 reported 'a lot less hard than it seems when going through the instructions.' Others reported more specifics such as Control5 who reported '*It is a lot easier to do the OxMET compared to MET-Home as it just involves sitting in one place and being assessed not faffing about. It is easier to know where items are on the OxMET because shops are obvious.*' Many comments referred to the interest and fun participants had when completing the METs (38 references). Most participants used synonyms of the word 'fun' and 'interesting.' Nineteen times the MET-Home was referenced as fun and interesting, and this was the same for the OxMET. Control26 referred to the MET-Home as a '*bit of a laugh.*' Control27 reported that the MET-Home was '*more interesting than other tasks I have done for research*', this was echoed by Stroke36 who said the MET-Home was '*more interesting than other tests.*' Control8 '*liked the idea of playing a game as an assessment.*' referring to the OxMET.

There were also negatively valenced comments, however, mostly related to the MET-Home. A third of participants made comments regarding the requirement of the MET-Home to call a real person; they expressed resistance to it due to wasting time (e.g., Control23 '*[I] don't want to waste time of membership services by calling*' (MET-home revised) and phone credit to complete this task (e.g., Control39 reported that they were '*unsure with the call as [I] didn't want to use credit to make the call unnecessarily*'). Additionally, feedback suggested multiple factors that affected performance on the MET-Home, including low mobility (e.g., Stroke24 reported that they '*worried about mobility during the task and going up and down stairs.*') and vision (e.g., Stroke5 reported '*[I] struggled with vision, seeing the instructions sheet.*') and resistance to being observed when completing the test (e.g., Stroke7 reported that '*. . . [I] didn't see the point in being observed doing it when I know I could do it.*').

Additionally, a small number of participants found the MET-Home stressful and intrusive (e.g., Control44 reported that they found the MET-Home '*challenging and intrusive. [I] felt that as the tasks expect you to go into private rooms and to know about medication taken, that it is intrusive*'). Many participants felt the MET-Home was stressful (e.g., Stroke4 reported that the MET-Home was '*Stressful. I feel uptight about not entering a room more than once etc. and keeping rules in mind.*' Stroke17 reported that it was '*stressful when [I] couldn't use the website to get membership cost as the website didn't state it.*' Finally, Stroke45 reported that it was '*infuriating not being able to go back into a room already visited.*' echoing Stroke4.

One participant remarked about the truncated questions on the OxMET: Control17 reported that they '*thought shortened questions [on the OxMET] were difficult to understand.*' Control17's comment referred to the display of the questions on the OxMET shopping list, as they are truncated to save space and only fully visible once selected.

Association of MET accuracy to qualitative feedback given I finally examined whether performance on either MET influenced valence of feedback given, that is, whether those with better performance were more likely to report positive

feedback or negative feedback. I ran a last simple logistic regression model, predicting valence of comments from accuracy (1 for positive valence, and 0 for negative valance – no feedback or neutral comments that it was completed were not included, but phrases like 'fine' were coded as positive – full data available in data repository for transparency), per MET task. For the MET-Home, I found that accuracy was not a predictor of valence of feedback (Tjur's R^2 = .05, OR=1.20, p=.07), and for the OxMET, accuracy was not a predictor of positively or negatively valenced feedback (Tjur's R^2= .02, OR=.68, p=.26).

3.1.5 Discussion

In the current study I aimed to compare the Oxford Digital Multiple Errands Test (OxMET), a computer tablet app, to the in-person Multiple Errands Test-Home (MET-Home). I examined their convergence and differences in feasibility and acceptability. I used a basic qualitative description approach to describe feedback from participants for both tasks.

When accounting for sample differences in education, I only found differences in performance on the OxMET not MET-Home. Identically to Alderman et al. [125], I found differences in OxMET total errors, task failures (termed omissions in OxMET) and frequency of errors when controlling for covariates, but no other metrics. In this way, in comparison to such a validated measure as the hospital version of the MET [125], the OxMET appeared to perform well, and the MET-Home did not. To be satisfactory, more differences would be present between healthy controls and stroke survivors, as most all MET research has found differences (see example, Lai, Yan, and Yu [132]). In the original psychometric investigation of the OxMET multiple group differences were found, but covariates of interest were not included. As this data was available to me (and on OSF; www.doi.org/10.17605/OSF.IO/8SUT5), I conducted secondary analyses on the participants from Webb et al. [130]. Using the original data and running ANCOVA analyses with education as a covariate of interest, significant differences were found on all measures (p<.001 for all analyses).

I re-ran the analyses including age as a variable as well, and found that the only non-significant results were in number of perseverations and number of omissions. Thus, when controlling for age and education, the only result that changed in the original investigation was group differences in perseveration which became non-significant.

In Burns et al. [126] they did not account for age and education effects and found multiple group differences. I was unable to check for whether age and education remove the group differences or not in their data. Moreover, the original MET-Home study used participants who were much younger than the current sample but well matched within their study apart from socioeconomic status and race. It is possible even in a well-matched study accounting for age and education may attenuate differences due to the known association of age and education with cognition [50].

I found that the OxMET and MET-Home were moderately associated, with most metrics from either test correlating with each other at least below an alpha level of .05, but many below the corrected alpha level too. Notably, most convergence was above the pre-defined alpha level of r=.30 [50, 129, 130]. The exception being MET-Home partial omissions and inefficiencies which did not related to OxMET metrics. This is likely due to the limited available variance, with the MET-Home comprising more complex tasks and more potential partial errors than in the shorter OxMET. Overall, however, the convergent validity was above benchmark on the core metrics of accuracy and errors, present despite the differences in length of administration and higher level of difficulty for the MET-home. I acknowledge, however, that convergence was not exceptionally high meaning there is still substantial unexplained variance in data. Further explanations could be the modality of assessment. Moreover, due to there being some missing data for MET-Home (not all participants completed the MET-Home), I am not capturing the full extent of the OxMET to MET-Home association. There may be additional influences in performance from those who did not participate that may have changed relationships found.

In addition, the way scores are calculated may affect results. For instance, the OxMET is objectively marked with clear criteria in the code for scoring, and there is no element of subjective interpretation. For example, partial omissions are

scored by how many times a participant entered a target shop and did not buy the correct target item, and/or how many times they entered a target question shop and did not answer the question, and /or how many times they answered a question without entering the target question shop. Whereas in the MET-Home, the 'partial completion' metric brings much subjective opinion, as it requires the examiner to judge whether someone has completed a task accurately or partly, without explicit guidance on how complete a task must be to be considered complete or partially completed for any potential example. Nevertheless, the MET-Home inter-rater reliability was previously shown to range from .88 to .96 [126]. In the limited research that has explicitly compared two METs, high associations have been found. One study created a revised version of the Baycrest MET (BMET-R) and found moderate to high correlations between BMET and BMET-R metrics [237], though similar to the present study, Clark et al. [237] did not report any significant associations using partial omissions. One other study using a virtual mall version of the MET, compared two healthy control groups (i.e., young and older participants) to a very small cohort of stroke survivors (n=9), and when using a divided sample of 29 participants (nine stroke and 20 older adults), found strong correlations between the virtual and in-person MET tasks, particularly for number of mistakes, partial task completions and for being inefficient. In both studies, sample sizes were too low to have statistical power greater than 80% thus reducing ability to make inferences on statistical significance. In addition, the high correlations may be explained by the small sample sizes (and thus large confidence intervals – not reported) which can lead to unreliably big correlations [30, 238].

With regards to feasibility, the majority of participants could complete both Multiple Errands Tasks. Where the MET-home was not attempted, this was mostly due to low levels of mobility, with some due to poor vision. I did not find any significant independent predictors of completion of either the MET-Home or the OxMET, due to inability to conduct this analysis.

In contrast to completion, the performance on the Multiple Errands tasks was differentially related to many individual factors. This suggests multiple factors

may affect performance, and this may have ultimate implications with regards to potentially adjusting future cut offs for impairment taking into account these factors (though I note the MET-Home currently has no formal normative data and cut-off scores – normative scores are in progress). In contrast, the OxMET score was not predicted by external factors other than overall 'global' cognitive functioning. Currently the OxMET has adjusted cut offs for age due to original findings that age was associated with performance in a large healthy control sample [130], though this was not in the context of other factors (e.g., it was a sole correlate).

In the brief examination of health economics I found there to be a higher upfront cost of running the OxMET due to the price of a tablet. If a computer tablet were already accessible, the OxMET quickly overtook the MET-Home in terms of low costs and staff time, especially for scoring. The MET-Home took longer than the OxMET and required manual scoring which incurred additional time. Moreover, the MET-Home had rare, but high, safety risk costs associated with potential falls, which were a particular risk of older age populations. It is important to acknowledge that examiners may explicitly wish to include the motor component in the traditional MET (e.g. due to the known effect of mobility on cognition) [13]. Moreover, in those who are mobile, but who complete OxMET, assessors may not get a true reflection of behaviour in everyday life, when removing this additional load. However, the goal of OxMET fundamentally is to assess executive functioning abilities which may affect everyday life, rather than to directly capture everyday life behaviors. Future research could investigate whether the additional motor component of the MET-Home would make it a stronger predictor.

Furthermore, whilst the OxMET is a brief screen for executive dysfunction that may reflect everyday behaviours, the MET-Home provides an in-depth assessment of functional behaviours in real life. The relevance of time costs depends on clinician time and setting. In a community setting with extra time in a 1:1 session, the MET-Home time demands may be achievable, but in in-hospital settings, where time is more constrained, the brief OxMET may be more appropriate.

I finally examined the feedback from participants on both tasks using a basic qualitative description approach and thematic analysis. Largely I found that there were more negative comments regarding the MET-Home, as opposed to the OxMET, especially surrounding the stress of the MET-Home. I interpret the findings of the qualitative examination of feedback to reflect the acceptability of the METs. I infer that there is less acceptability for the MET-Home as it can be stressful and not appropriate for those with low mobility. I take this interpretation from over 21 references to not wanting to complete item one of the MET-Home. Note, I discussed item one with the MET-Home authors, who asserted that this item cannot be removed or replaced with a non-call option. I infer that as the OxMET was seen as simple and straightforward, and in some cases fun, with rare negative comments, that it is more acceptable to healthy adults and stroke survivors. The OxMET may be too simple for more independent individuals, and the MET-Home too taxing and risky for less mobile or more disabled individuals. I note that the feedback presented here, is only relevant in the comparison of OxMET to MET-Home, and feedback may differ when comparing to other METs such as virtual versions. Where a clinician is choosing between a physical MET or a computerized MET, participant acceptability must be factored in. For those who may have mobility or other issues with the stress of a physical MET, the OxMET, or indeed other computerized METs may be preferred.

We also acknowledge that many comments made were neutral or positive, meaning that the majority of participants did not have negative issues with either MET.

The participant feedback matched up to the quantitative results in one major way, that there were greater barriers found to completing the MET-Home than the OxMET. Our quantitative results indicated mobility as a predictor of MET-Home success, and this was matched in participant feedback. The lack of negative comments and few comments regarding vision affecting completion, matches up with the high rates of OxMET attempts/completion.

Limitations of the study

While I was statistically powered to detect the benchmark correlation of r>=.30, I added justifiable and important covariates of interest which may have reduced the statistical power to detect effects. I was one participant away from the minimum sample size required. Additionally, only 48 participants were stroke survivors, limiting the generalizability to stroke survivors. I collected both healthy controls and stroke survivors to contribute to the normative data base for the MET-Home and OxMET. Our method of qualitative data collection had limitations. I collected feedback from participants after completing each MET, and the feedback was not recorded verbatim for all. I therefore describe the feedback only descriptively and did not conduct a full qualitative analysis. Initially, I recorded feedback that I viewed as the same in a standard way to aid in analysis. For instance, those that refused item one of the MET-Home were coded as the same phase 'did not want to complete item one of the MET-Home', but exact wording was lost. I realized the importance of collecting verbatim information during the project and adjusted the methods to do so for the remainder of the project. As such, some meaning may have been lost in the records, though I believe the core feedback was retained.

Future Research

In light of the current study, there are several avenues for future research. One is to compare the OxMET to other computerized assessments of executive functioning. For instance, it would be good to compare OxMET to the Virtual Multiple Errands Test [140] or Virtual Mall [136], both of which include shopping scenarios which the MET-Home does not. A direct comparison of OxMET to computerized METs would further establish the specific acceptability and feasibility of OxMET as a tool. In addition, comparison of OxMET to non-MET computerized assessments such as the Computerised Multiple Elements Test [239] or Computerised Wisconsin Card Sorting Test [240], would provide further evidence of feasibility, acceptability, and convergence.

Another important question would be to include the perspective of the clinicians administering a MET or OxMET. Research should examine in particular if clinicians feel there is any added benefits of OxMET to their own practice.

3.1.6 Conclusion

The study highlights good convergent validity for the digital OxMET with the in-person MET-Home. The OxMET digital administration provides a more acceptable and inclusive assessment, especially to people with mobility restrictions and more severe stroke. The MET-Home may be more suitable to mild stroke and those with lower levels of physical disability. The OxMET is not intended to replace in-person direct observation which may provide additional valuable information to clinicians.

3.2 Predictive validity of the Oxford Digital Multiple Errands Test (OxMET) for functional outcomes after stroke.

This section is partially altered from the peer-reviewed publication:

***Webb, S.S**. & Demeyere, N.(2023). Predictive validity of the Oxford digital multiple errands test (OxMET) for functional outcomes after stroke, Neuropsychological Rehabilitation, DOI: 10.1080/09602011.2023.2247152*

3.2.1 Abstract

3.2.2 Introduction

Building on the work of establishing the OxMET as a standalone screen for executive dysfucntion [130] and the just presented work of converging the OxMET with real life behaviours measured through another MET, the predictive validity of the OxMET was assessed. The current section differs in that the previous section examined the *verisimilitude* of assessments, or the degree to which the test reflects the demands of real life, and the current section measures the *veridicality*, or the degree to which I can make predictions or inferences about real life on the basis

of test performance [241]. Widely used tests for executive function, like the trail making test [242] (TMT), were not designed to reflect real life and therefore lack verisimilitude. Nevertheless, these tests often show veridicality in their association to activities of daily life (e.g., Lipskaya-Velikovsky et al. [243]).

Predictive validity is a form of criterion validity where one evaluates the magnitude of an association between the current test and a criterion test at a later date [21]. For example, Katz, Tadmor, Felzen, and Hartman-Maeir [244] examined the predictive validity of the Behavioural Assessment of Dysexecutive Syndrome (BADS) [139] in predicting functional outcomes for those with schizophrenia, in an approach similar to the current investigation. Whilst the BADS is one of the most popular cognitive assessments of executive functioning, it takes approximately 40 minutes to administer, and requires the administrator to have specific qualifications and training (defined at level 'A' or 'B' qualifications from the Pearson website; https://www.pearsonclinical.co.uk/). Whilst the DEX questionnaire aims to reflect everyday problems and experiences, the battery as a whole does not directly relate performance to a real-life scenario.

In a systematic review of the MET literature, Rotenberg et al. [129] found there were 12 studies that examined the relation of the MET to everyday behaviours with 24 hypotheses from these papers [129]. There were inconsistent relationships between METs and everyday behaviours, with 11 of 24 hypotheses being confirmed [129], and mixed evidence being found even when using the same measure across studies. The most frequently used measure to reflect real life problems was the dysexecutive questionnaire (DEX) taken from the BADS, and many MET studies have not reported consistent associations of their METs with the DEX [129]. However, few MET studies looked at the relation to activities of daily living explicitly through assessments directly assessing abilities, thus whilst the face validity of the MET is extremely good, its predictive validity for real life outcomes remains unclear.

The current study provides evidence to determine the predictive validity of the OxMET with regards to meaningful functional outcomes both in a subacute stroke rehabilitation setting and longer-term in stroke recovery (6-months post-stroke).

The association between the performance of a stroke sample on OxMET and routine tests of functional ability collected from a stroke rehabilitation unit were examined. A consecutive proportion of the sample were followed up to examine predictive abilities of the OxMET taken sub acutely to activities of daily living 6 months later.

This study was pre-registered, and data and analysis scripts to recreate the manuscript are openly available in CC-BY 4.0 license (doi.org/10.17605/OSF.IO/WEP3V).

3.2.3 Methods

The method of determining the sample size was reported, any data exclusions were noted (if there were any), all manipulations were described, and all measures used in the study were documented (Solution for transparency by Simmons et al. [171]). The manuscript adheres to the COSMIN guideline for studies on measurement properties [217] and the STROBE cohort checklist [188]. This study protocol was approved by the National Research Ethics Committee South Central – Oxford C Research Ethics Committee (OCS-RECOVERY; REC reference: 18/SC/0044, IRAS project ID: 241571). All participants provided written or witnessed informed consent.

Pre-registration

The pre-registration for time point one (participants in rehabilitation unit) involved correlating OxMET scores with admission and discharge functional outcome measures (pre-registration `https://osf.io/z37h8`). Sample size was determined a priori by the smallest effect size of interest which was a correlation of r=.30, a power of 80%, and an alpha of .05, which lead to a need for a minimum of 66 individuals. The smallest effect size of interest was determined due to being a correlation benchmark for convergence [129]. The aim was to recruit as many participants as possible at time point one (recruitment in hospital) until a minimum of 66 participants was reached at follow up, to account for expected attrition (previous inclusive recruitment of moderate stroke has shown up to 50% attrition, e.g. see Demeyere & Williams, et al. [245]).

I deviated from the pre-registration variables but not hypotheses: I only correlated outcome measures taken closest to administration of the OxMET rather than specifically at discharge or admission to ensure complete data. I added disability outcome measures to examine different aspects of function that were clinically relevant. I added in an additional analyses with Montreal Cognitive Assessment (MoCA) data - where available - to investigate the specificity of the relationship between activities of daily living with the OxMET in comparison to a global cognitive screen.

Measures

Multiple functional outcome measures from basic independence to more holistic functioning including well-being were included based on clinical recommendation from the stroke rehabilitation unit. These measures were collected in addition to the OxMET.

Functional Outcome measures collected at time point 1 (baseline) The Barthel Index was included as a commonly used and validated tool to classify basic independence [222, 223]. The stroke specific Therapy Outcomes Measure [246] (TOMS) was included as a more holistic measure of functioning, on recommendations of Occupational Therapists at the unit. The TOMS examines four domains each rated by a physio- or occupational-therapist: Impairment (i.e., physical, cognitive, sensory impairment etc), Activity (i.e., how able or independent one is at purposeful action), Participation (i.e., ability to involve oneself in their own life, be that socially, emotionally, educationally etc), and Well-being / Distress (i.e., scales from uncontrollably distressed/emotional to appropriate levels of well-adjusted emotionally) (Enderby et al., [246], pp. 207–209). Each domain is scored on a 11-point scale (0 being the worst and 5 being the best, with half-points used to indicate slight deviation from the integer categories) (Enderby et al., [246], p. 23). I created a summed total score of the TOMs as an outcome measure to compare to OxMET performance and reduce multiple comparisons.

The modified Rankin Scale (mRS) and The National Institute of Health Stroke Scale (NIHSS) were collected, detailed previously.

Functional Outcome measures collected at time point 2 (follow up) Participants who completed follow-up, completed cognitive screening with the Oxford Cognitive Screen [38] and OCS-Plus [50] and the Montreal Cognitive Assessment (MoCA) [37], and were administered a series of questionnaires as part of the OCS-RECOVERY protocol. In the current study, only the Nottingham Extended Activities of Daily Living (NEADL) [247] 66pt scale [225], the MoCA, and the mRS for disability and dependence level were included as outcome measures.

Participants

A total of 117 participants were recruited to this study, with 66 stroke survivors completing the 6 months follow up. Attrition until follow up occurred due to: loss of contact (n=19), death before discharge (n=2), death before follow up (n=10), no interest in completing the follow up (n=7), too unwell to complete follow up within 2 month time frame (n=6), and that I reached 66 participants before their follow up was due (n=7). Participants were recruited and followed up between November 2020 and December 2022. No eligibility screening data were recorded. Inclusion and exclusion criteria for stroke survivors were the same as reported in chapter 1, chapter 2, and in the previous section.

The summary demographics of the sample are presented in Table 3.5. Missing NIHSS, pre-morbid mRS, and mRS at follow-up values were imputed using the mean for each variable respectively; the proportion of missingness is shown in Table 3.5. The distributions for each of the functional outcome measures are displayed in Figure 3.2 for context of the sample. It is shown in Figure 3.2 that participants were moderately disabled on the mRS, had moderate strokes on the NIHSS, ranged mildly on the Barthel Index, and a wide range of holistic functioning scores was observed on the TOMS. The range in participants' abilities is described, reflecting a realistic representation of an in-patient rehabilitation setting.

Table 3.5: Summary of sample characteristics for all participants and those followed up 6 months post-original consent

	All participants		Followed up	
Characteristic	N (% missing)	Values	N (% missing)	Values
Age (M(SD))	117 (0%)	74.44 (12.88)	66 (0%)	73.94 (12.68)
Education (M(SD))	117 (0%)	13.13 (3.25)	66 (0%)	13.52 (3.47)
Handedness	117 (0%)	R: 88.03%; L: 11.97%;	66 (0%)	R: 81.82%; L: 18.18%;
Sex	117 (0%)	M: 56.41%; F: 43.59%;	66 (0%)	M: 60.61%; F: 39.39%;
Ethnicity	117 (0%)	White-British: 89.74%; White-Other: 6.84%; Black African: 0.85%; Black-African: 0.85%; Black-Carribbean: 0.85%; Other-Asian: 0.85%;	66 (0%)	White-British: 92.42%; White-Other: 4.55%; Black African: 1.52%; Black-Carribbean: 1.52%;
NIHSS (M(SD))	117 (0%)	8.32 (5-11)	66 (0%)	8 (4-11)
Time Since Stroke (M(SD))	106 (9%)	26.18 (25.16)	64 (3%)	201.76 (26.82)
Stroke type	117 (0%)	ischaemic: 80.34%; intracerebral haemorrhage: 14.53%; ischaemic with haemorrhagic transformation: 4.27%; subarachnoid haemorrhage: 0.85%;	66 (0%)	ischaemic: 74.24%; intracerebral haemorrhage: 19.7%; ischaemic with haemorrhagic transformation: 4.55%; subarachnoid haemorrhage: 1.52%;
Stroke side	117 (0%)	R: 52.99%; L: 29.91%; B: 17.09%;	66 (0%)	R: 53.03%; L: 30.3%; B: 16.67%;

Bamford Stroke Classifification	116(1%)	PACI: 37.93%; LACI: 29.31%; PACH: 8.62%; TACI: 7.76%; LACH: 5.17%; POCI: 3.45%; LACS: 2.59%; PACS: 2.59%; POCH: 0.86%; TACH: 0.86%; TACS: 0.86%;	65(2%)	PACI: 35.38%; LACI: 29.23%; PACH: 10.77%; LACH: 7.69%; TACI: 6.15%; POCI: 3.08%; LACS: 1.54%; PACS: 1.54%; POCH: 1.54%; TACH: 1.54%; TACS: 1.54%;
Language Disturbance	116(1%)	None: 60.34%; Dysarthria: 17.24%; Expressive Aphasia: 13.79%; Mixed Dysphasia (Expressive + Receptive): 6.9%; undetermined: 1.72%;	65(2%)	None: 66.15%; Dysarthria: 15.38%; Expressive Aphasia: 10.77%; Mixed Dysphasia (Expressive + Receptive): 6.15%; undetermined: 1.54%;

Procedure

Participants were approached by myself at bedside in the rehabilitation unit after discussion from the MDT about inclusion/exclusion criteria. Following informed consent, participants were screened with the OCS and if there was enough time before other therapy or discharge, they were administered the OxMET. The Barthel, TOMS, NIHSS, mRS, and sometimes the OCS, were administered by the clinical team separately. Clinical and some demographic information were collected from medical notes. Participants were followed up 6 months later in person at home. Questionnaires were sent in advance via post.

Data analysis

As pre-registered, I examined correlations between functional measures taken in the rehab setting to OxMET metrics through Spearman's Rho correlation to establish any relationships. Next, I looked at the predictive abilities of OxMET performance in the rehab setting to activities of daily living 6-months following study participation. A regression analysis of the OxMET metrics and outcomes of the NEADL total score was conducted, with covariates of change in mRS score, NIHSS, age, and education. The study was powered to run the regression analysis on a minimum of 66 participants. I conducted these regression models to investigate the potential predictive ability of OxMET accuracy. The OxMET accuracy score was chosen as it is the main outcome metric and intended to simplify impairment classifications and interpretation. As I only collected other cognitive data at 6-months, I compared the prediction of the NEADL score by OxMET score, to the prediction of NEADL score by 6-month MoCA score, using the same covariates. I then included the 6-month MoCA in the baseline OxMET prediction model of 6-month NEADL to assess the independent contributions of both measures. For correlations and regressions I used a convergence benchmark of r=.30, with acceptability of correlations above .19 as per other very commonly used neuropsychological batteries (see Demeyere, et al., [50] which discussed convergence acceptability).

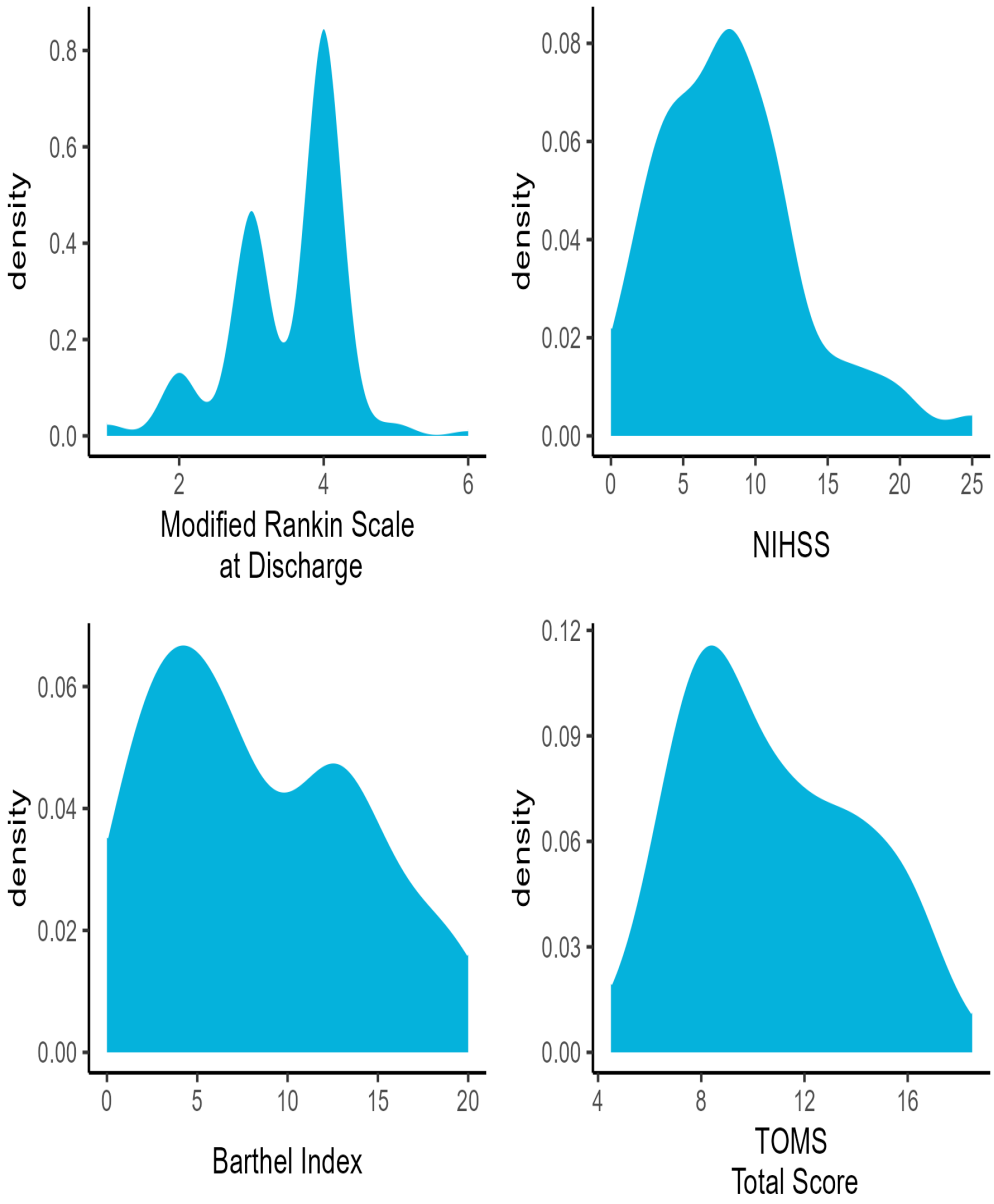


Figure 3.2: Illustrates the density distributions of functional outcome measures from the stroke sample (<3 months post-stroke).

Data wrangling and statitical software

Most data wrangling was completed in MATLAB [190] as the executable file for the research original OxMET produced MATLAB structures for performance data. All statistical analysis and some data sorting was computed in R studio [153] version 4.2.2. I used the following packages for the production of the RMarkdown manuscript and analysis: *bookdown* version 0.32 [196]; *tidyverse* version 1.3.2 [248]; *readxl* version 1.4.2 [198]; *cowplot* version 1.1.1 [249]; *ggplot2* version 3.4.0 [201]; *kableExtra* version 1.3.4 [204]; *httr* version 1.4.5 [250]; *ggsignif* version 0.6.4 [251]; *Hmisc* version 4.7-2 [252]; *pROC* version 1.18.0 [199]; *data.table* version [253]; and *gridExtra* version 2.3 [254].

3.2.4 Results

Associations with baseline outcome measures

Accuracy, the main outcome metric from the OxMET, was significantly related to the TOMS total score ($r(97)$=.27, p=.007) and mRS (r(114)=-.30, p=.001) below a multiple comparison corrected alpha (p< .05/3 comparisons), and the Barthel Index ($r(77)$=.23, p=.04). The correlation coefficient for the mRS was at the benchmark for convergent correlation, and the Barthel Index and TOMS total score was within the acceptable correlation criteria. Correlations between secondary OxMET metrics and the functional measures are presented in Table 7.6 in the appendix.

Predicting longer-term functional outcomes

Baseline OxMET accuracy was a significant predictor of NEADL total score (R^2 adjusted = .29, $F(5,57)$= 5.98, p<.001, B_1 = 2.1, p<.001), when accounting for variance explained by change in disability and dependence (change in mRS score), age, education, and stroke severity at baseline (NIHSS score), and with a correction for multiple comparisons at the level of .05/4 (p< .013). I found that OxMET accuracy demonstrated a clear positive predictive association with the NEADL. The simple associations between OxMET accuracy and NEADL scores are presented in Figure 3.3.

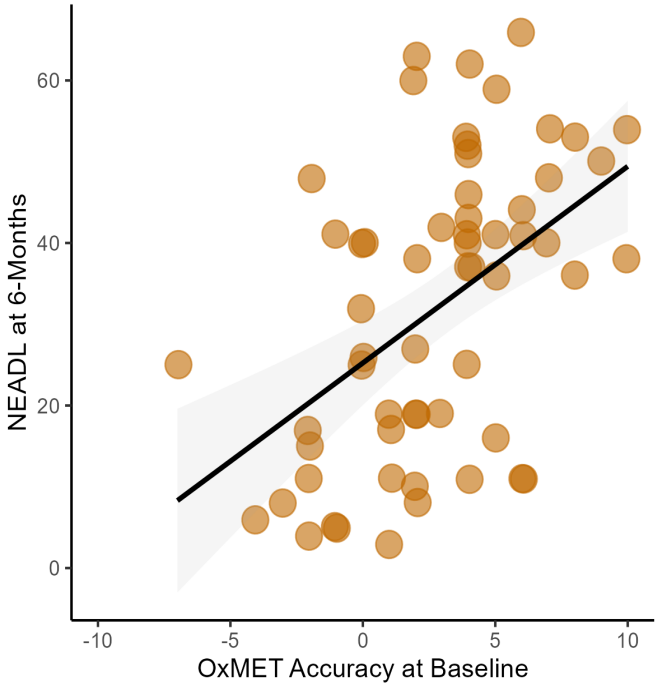


Figure 3.3: Visually jittered scatter plots between OxMET Accuracy scores and functional outcome measures taken at 6-month follow up.

Finally, a linear regression with NEADL score predicted by MoCA total score taken at 6-months was completed using the same covariates and I found no significant association with MoCA, only with stroke severity (R^2adjusted = .21, F(5,43)= 3.535, p=.009, B_{MoCA} = 0.7, p=.06). When I included the 6-month MoCA in the OxMET prediction model, only the OxMET accuracy metric was a significant predictor of NEADL score at 6-months (R^2adjusted = .34, F(6,42)= 5.03, p<.001, B_{OxMET} = 1.88, p=.004, B_{MoCA} = 0.44, p=.251). MoCA was only available for 50 out of 66 participants, and the differing sample sizes for the models will have affected the power to detect significant effects. However, I also note that the strength of the association increased for the OxMET with the second model including MoCA.

3.2.5 Discussion

I aimed to provide evidence for the predictive validity of the Oxford Digital Multiple Errands Test (OxMET) using a subacute (<2 months post-stroke) stroke survivor sample seen at two time points (baseline in a stroke rehabilitation unit and then

6 months later at home). The main analysis of the current study concerned associations between subacute OxMET and functional abilities at baseline and at 6-month follow up. Here I further evidence the case for usefulness of the OxMET by demonstrating predictive validity.

The OxMET significantly associated with the Therapy Outcomes Measure total score (TOMS) [246] and modified Rankin Scale (mRS) [228], but not the Barthel Index [223], when correcting for multiple comparisons. However, the Pearson's r did not breach the threshold for convergence (r>=.30) for the TOMS, and as such I can only cautiously interpret this association. For 6-month follow up, I found a moderately significant association between OxMET accuracy and Nottingham Extended Activities of Daily Living (NEADL) scale [247], over and above variance explained by stroke severity, demographics, and disability change. These findings suggest the OxMET, as an executive functioning screening tool, can be used as a valid assessment with a good degree of predictive validity to functional outcomes in activities of daily living. Nevertheless, inferences about real life functioning should still not be made on an individual prediction level.

Furthermore, I found a significant predictive relationship between OxMET and 6-month NEADL score, but I did not find one for concurrent MoCA score and NEADL. It remains possible that there could have been a predictive relationship from baseline MoCA to future functioning, but I did not collect baseline MoCA scores here. Indeed others have found such a relationship [255, 256]. However, here I found a specific relationship from baseline OxMET, which could not be explained by concurrent MoCA scores.

There is limited research regarding the association between variants of the Multiple Errands Test (MET) and basic functional outcome measures relevant for rehabilitation settings. The Barthel index score is commonly used as an exclusion/inclusion criterion rather than as an outcome measure due the physical nature of the MET (e.g., Burns et al., [126]; Morrison et al., [257]; Raspelli et al., [137]). As such it is difficult to compare the OxMET and other METs in their relation to basic functional outcome measures. However, other more complex

functional measures have been investigated in comparison to METs. For example, Dawson et al.[133] compared their Baycrest MET version to assessments of motor and cognitive impairments and complex activities of daily living in stroke survivors who were younger, more educated, and more chronic (8.6years post-stroke) than the current sample [133]. They found significant moderate to large correlations above .30 between their error scores, but not accuracy. Maeir et al. [258] investigated the associations of errors in the hospital version of the MET [124] with participation in the community 3-months post-stroke [258]. The study found correlations above .30 for most MET metrics except task failures, and they did not use an accuracy metric. When I re-ran regressions including the error OxMET metrics, I found that accuracy was still the only metric related to activities of daily living, over and above change in mRS, NIHSS, age, and education. This strengthens the justification for using an overall outcome metric, to make clinical interpretation easier. Differences between results may be due to the current sample sample, which was more severely affected by their stroke, older, and with greater levels of disability.

Study limitations

There are limited tools which can measure functional ability in a standardised and psychometrically validated way in in-patient settings. I was reliant therefore on functional assessments which revolve around basic ADLs, such as the Barthel, or assessments which are confounded by physical impairments. This is a limitation in that these tasks cannot reflect real life behaviours which are complex. However, the Barthel and TOMS are commonly used in clinical practice and provide an easily interpretable scale to compare to the relatively new OxMET task.

Separately, my study deviated from the standardised administration of the OxMET in that I used supplemental instructions developed with the speech and language team to aid comprehension of instructions for participants with lower communication abilities. The use of images paired with broken down and spaced-out text has now been implemented in the latest version of the OxMET app available on the appstores. The small differences in administration could have conceivable led

to higher performance, though arguably also allow for the interpretation of a more specific executive deficit, rather than a potential confound from a problem in complex instruction comprehension. There was no evidence of an overall better performance though, with the current stroke cohort as a whole scoring well below the healthy ageing data [130]. Updated normative data will continue to be added to the app and I will continue to provide open data for readers interested in direct comparison.

I recruited only from a specialist stroke in-patient rehabilitation unit and not from acute or hyper acute, or out-patient settings, restricting generalizability. I feel that this is countered by using a moderate stroke sample, where most research uses mild stroke cohorts.

Future Directions

The OxMET has been psychometrically validated as an executive functioning screening tool, and now has evidence to support its predictive validity. Future acceptability and feasibility of the OxMET in clinical practice, administered by allied health professionals, will need to be formally assessed.

3.2.6 Conclusion

The OxMET total score associated with basic functional measures and predicted activities of daily living. Its associations with clinically relevant and activity relevant measures provide key evidence to support its use in real-world clinical settings.

Discussion

Throughout this thesis, I have presented works regarding different cognitive impairment screening tools, with the aim of providing validity evidence for their use in different contexts. Part one of empirical chapter one addressed the impact of the COVID-19 pandemic on cognitive assessment and found that the modality of assessments changed from in-person to remote, but the frequency of assessments did not necessarily decrease. Part two of the same chapter detailed a small data set providing initial validity evidence of the use of the Tele-OCS, a remote version of the OCS, with chronic stroke survivors. The Tele-OCS was found to have satisfactory associations with the in-person OCS. In empirical chapter 2, I provided validity evidence for the use of the OCS-Plus, a tablet-based cognitive screen aimed to detect milder domain-general cognitive impairment (particularly in Memory and Executive Function), in both subacute (<3 months post-stroke) and chronic (~6 months or greater post-stroke) stroke. Empirical chapter 2 revealed that the OCS-Plus had satisfactory construct validity and evidence was provided for its sensitivity/specificity. Empirical chapter 3 concerned the Oxford Digital Multiple Errands Test (OxMET) and was divided into two parts, with the first part providing validity evidence of the OxMET with regards to associations with existing Multiple Errands Tests, the second part with regards to predicting future functional outcomes.

This discussion will largely follow the structure of the introduction, focusing on assessment of primary/gross level cognitive impairments, subtle cognitive impairments, and ecological validity of cognitive assessments. Findings will be summarised in relation to the Standards [21] for validity evidence as well as integrated towards future research directions.

3.3 Assessment of primary/gross level cognitive impairments post-stroke

Validity evidence for cognitive screening tools is paramount in accurately flagging a given stroke survivor with probable cognitive impairments for further in-depth assessment if necessary. With regards to the Standards of Educational and Psychological Testing [21], ensuring that a test is designed for a given population, that it accurately measures constructs it claims to, and that it is used in validated applications, is crucial to ensuring integrity of the test and accurate interpretation.

In response to the COVID-19 pandemic, I conducted a research survey to estimate the impact the pandemic had on cognitive assessments. This was a small scale UK-focused research study due to the time sensitivity of the subject matter. The survey required respondents to state the modality and frequency of cognitive assessments pre-March 2020 and post-March 2020. I found the frequency of assessments did not change after the pandemic began, but the modality shifted, with more remote assessments. Cognitive assessment can be essential to highlight impairments which may have impacts on future recovery and functional outcomes [2, 3, 123, 213, 255], so not reducing the number of cognitive assessments means stroke survivors were, in principle, having impairments highlighted at the same rates as pre-pandemic. This is in contrast to wider health systems struggling with provision of care.

One population-based study examined the impact of the pandemic on delivery of radiotherapy for cancer in the NHS in the UK [259]. The authors found that in general, provision of radiotherapy declined in April 2020 compared to previous years, yet provision of fractionated (sessions split up) radiotherapy actually increased for some cancer diagnoses [259]. Regarding subacute stroke care, during the pandemic there were far fewer stroke admissions than previous years, yet there were found to be no differences in rates of thrombectomy provision, or time, and no difference in onset of stroke time to hospital admission [260]. In the context of empirical chapter 1, part 1, the differences in rates of admissions or provision of care were differentially affected during the pandemic. Finding that the frequency of conducting

cognitive assessments was not affected by the pandemic, highlights the quick and adaptive response of healthcare workers to use of remote assessments to ensure they provide the best care possible for stroke survivors. It must be noted that the numbers did decline, but that this was not a significant decline, which may reflect the overall lower stroke admissions in 2020 [261, 262]. In the pandemic in-patient hospital services were not as restricted. Once in a hospital, care was taken to avoid transmission of the disease but stroke rehabilitation, for example, was still being undertaken, as well as in-patient cognitive assessment [263]. The shift in modality did raise concern regarding the validity of tools used for cognitive assessment and screening. As evidence in empirical chapter 1, part 1, many assessments including the Oxford Cognitive Screen [38] were reported to be used without appropriate validation for remote administration. The OCS was designed to be used as cognitive screen in a hospital environment, adaptable for breaks in testing and for cardinal symptoms of stroke such as aphasia, apraxia, and visual field deficits [38]. Very little, if any to my knowledge, evidence has directly examined or discussed the impact of the COVID-19 pandemic on in-patient cognitive screening. One study examined the delivery of inpatient neuropsychological assessment in general (not specific to stroke) during the pandemic [264]. The authors generated a reduced contact neuropsychological battery on the basis that patients prefer in-person assessment over remote assessment [264, 265]. The test battery included tasks that could be administered without contact, but in the context of stroke, the tasks were largely verbal and thus unsuitable for most stroke survivors with communication impairments. The research study did not validate this new test battery for reduced contact in-person assessment, and much like other non-validated tests developed in the pandemic, is therefore not suitable for use based upon the Standards [21].

Following the initial response to the pandemic, there was a clear call (included in Webb et al [87] as well as multiple national guidelines [74, 164]) for immediate action to generate validity evidence for standardised tools. In part two of the first empirical chapter, I provided some validity evidence for the Tele-OCS, which was developed in the peak of the early COVID-19 pandemic in response to many

requests received by clinicians for a pragmatic adaptation to allow remote testing with the OCS. I detailed the development of the Tele-OCS, and provided data for 40 chronic stroke survivors. In terms of severity of cognitive impairment, this sample was mild, with on average 5 subtasks being impaired. Nevertheless, the cognitive impairments that were detected were captured in both modalities. The analyses on the small, but statistically powered sample, revealed that overall the modalities for the OCS were not too different, though some differences were noted in a small selection of subtasks (i.e. picture naming and executive functioning subtasks). In relation to existing literature, small deviations are expected in performance [183].

Nevertheless, whilst this initial study provided some validity evidence for the use of the Tele-OCS, the evidence is not convincing enough to promote usage of the Tele-OCS in practice currently without further evidence. This is for several reasons. The first is that Tele-OCS has no published evidence of use in a more impaired cohort. The current study used a mild sample, without severe cognitive impairment. It would have been ideal to use an acute sample with a wider range of cognitive-domain impairments. The range in impairment classifications, and thus range in continuous performance scores, would have allowed for better statistical power to determine reliability between modalities. The choice of using chronic stroke survivors does fit with its clinical use in community settings with chronic stroke, and a large proportion of chronic stroke survivors still have primary cognitive impairments later in recovery [9, 11]. The prevalence of primary impairments may not be as high, in comparison to prevalence acutely (though note, still approximately 68% had an impairment at 6 months in Milosevich et al., [11]), however, the Tele-OCS would most likely be used in the community, or with those who are in the more chronic stages of stroke. For services like early supportive discharge or community stroke therapy, there may be more use as these teams typically travel to assess stroke survivors. If a cognitive assessment is the only thing to be completed for a session, and a participant lives far away, the Tele-OCS may be suitable. In this case, the most severely affected may not be assessed with Tele-OCS anyway and so there is less concern about the validity. Nevertheless, more acute stroke survivors shortly

after discharge home (e.g., in Early Supported Discharge settings), will require assessments and Tele-OCS needs further evidence for its use in this population.

In summary, the pandemic significantly altered the way in which cognitive assessments are administered. Regarding assessment of primary cognitive impairments in inpatient settings, little research has examined this and perhaps this is due to the lack of change in inpatient care in the pandemic. Clinicians were using a version of the OCS remotely in non-inpatient settings, but this had not been validated. Some validity evidence was presented in empirical chapter 1, but much more detailed validity evidence is needed before Tele-OCS can be promoted for use in clinical practice.

3.4 Assessment of complex and/or subtle cognitive impairments post-stroke

Empirical chapter 2 presented validity evidence of an assessment designed to detect subtle or milder cognitive impairments in stroke. Here, for the first time, validity was demonstrated for the intended population, in contrast to the original psychometric validation and normative paper for the OCS-Plus [50]. The original norm and validate study used 320 relatively highly educated healthy adults from across the UK and Germany. In the current stroke sample (n=347), stroke survivors who were either within 3 months post-stroke or more chronic (>~6 months post-stroke) were tested. Like the original study, the OCS-Plus was compared to other screening subtasks and domains for convergence and divergence. A further departure in the current study was with regards to the comparison tasks chosen. The chosen comparison tasks in the original psychometric investigation were not necessarily the ideal choice, and some tasks were not best matched in format, but only construct measured. In the current investigation, each of the convergence tasks were better matched on construct as well as format to better remove attenuation of correlations due to confounds. In the original OCS-Plus study, healthy adults were at ceiling on some OCS-Plus tests thus precluding some statistics like reliability

from being conducted well. In the current study the use of stroke survivors and their heterogeneity automatically increased variance, and the mixture of time post-stroke sufficiently increased the between-sample variance, but kept within-sample variance small, leading to a higher powered and more reliable study [266]. The overall impression of this study was that the OCS-Plus is an appropriate tool for use in stroke, both closer to stroke date and more chronically, with good convergence and divergence across most subtasks. There were a small number of tasks which did not reach divergence criteria, however, the average effect size found in the convergent validity was .43, and for divergence it was .18 and the difference was statistically significant (Welch Two Sample t-test: $t(45)=6.24$, $p<.001$, 95% CI [.17-.34]).

When determining convergence and divergence of neuropsychological assessments, it is important to have a reference standard, ideally a ground truth reference (or 'gold standard') with which to compare to the validated tool. There is a larger question regarding whether any given test is the gold standard and thus whether individual correlations are the most appropriate method. Convergence of tests with others that measure the same construct is certainly an accepted and widely used methodology for construct validity evidence [37, 38, 50, 267–269]. In the original psychometric investigation of the OCS-Plus, consideration was raised as to the acceptable benchmark for convergence, with some previous studies accepting convergence above $r=.19$ [50]. There is generally no clear benchmark for convergence [267], and throughout my thesis I have used an acceptable benchmark used in MET literature as the cut off [129, 130, 270, 271] rather than no standardised cut off at all. When examining series of commonly used neuropsychological batteries, it is clear that the criterion of .30 is generally accepted. For instance, in an investigation of the Naming Test of the Neuropsychological Assessment Battery (NAB), the authors correlated the NAB with the Boston Naming Test [272], and found correlations around $r=.30$. Note, for this study, the authors claimed that convergence was found between the NAB and Boston Naming Test, yet found multiple correlations between the NAB and visual and verbal memory tests also [273]. In fact the Boston Naming Test has its own criticisms as a measure of naming

or language ability, yet is considered itself to be well-established and validated [274], and the reference standard. It is not the case that this is specific to language assessments. When examining the criterion and convergent validity of the Montreal Cognitive Assessment (MoCA) against a standardised neuropsychological battery, convergence was suggested for the domains of memory, visuospatial, language, memory, attention, and executive function, but multiple cross-correlations were found [275]. The authors claim these overlapping correlations (mostly all over the r=.30 benchmark) support a two-factor model of the MoCA, not the purported five domains being measured [275].

In the context of this empirical chapter, it is possible the results found were attenuated due to imperfect validity of comparison tasks. Where no gold standard or ground truth exists, efforts were made to ensure the highest quality format and construct matched convergence comparisons, so if no associations were found, then the OCS-Plus likely did not measure the purported construct. Yet, it is still possible, and evidenced, that correlations may not always be present for items with high face validity. For instance, Trails A accuracy (a combination of total correct connections of a baseline to connect squares descending in size, and then separately to connect circles ascending in size) was not related to CLQT symbol trail baseline (connection of circles ascending in size order), yet both on the face of it are measuring the ability to accurately connect shapes following a basic rule. Additionally, delayed recall from the OCS-Plus was not statistically related to MoCA delayed free recall, yet both are recalling five words after a similar delay in time, with words that are similar in frequency in the English language. Thus, in terms of whether two tasks appear to measure the same thing in the same way, there should have been correlations between these subtasks but there were not. Alternatively, sometimes the MoCA and OCS-Plus were administered in the same session, and sometimes they were not. It is possible, though not quantified by the data presented, that having the tasks administered consecutively could introduce confounds even where tests were counterbalanced. For instance, there are some floor effects in the MoCA delayed recall task present in different samples [276–278]

which may be introduced due to the time between MoCA encoding and recall which may be exacerbated by fatigue of this protocol.

In summary the OCS-Plus has previously been examined in a healthy control cohort and good convergence and divergence validity evidence was presented. In the current empirical chapter, the OCS-Plus was again examined in the context of construct validity but in two stroke cohorts, as well as adding information on the diagnostic accuracy of the OCS-Plus compared to a standardised neuropsychological battery. The OCS-Plus was again shown to meet criteria for construct validity and diagnostic accuracy, however, there are considerations about whether gold standards or ground truths are good enough to fully establish construct validity in any assessment.

3.5 Ecological validity of assessments of cognitive impairments post-stroke

In the final empirical chapter of my thesis, I again built upon my examination of construct validity in cognitive assessment, but with a focus on associations with external criteria. Specifically, the first empirical project regarded comparing the OxMET with the MET-Home to partially examine convergence, but to go further and examine feasibility and acceptability of the measures. In the second empirical project, I examined the ability of the OxMET to predict clinically relevant functional outcomes. Again, the MoCA [37] was used in this chapter, and although it is not a gold standard as discussed extensively in the introduction, the MoCA is widely used and was considered the best cognitive screen in stroke based on historical use [45, 49]. To compare and consider the MoCA in the context of newly developed tests, including the OCS-Plus, and now the OxMET, is to give a familiar reference to clinicians who know and use the MoCA and can thus contextualise found associations.

I chose the multiple errands test (MET) [106] as the primary focus of this chapter, due to the rich in-depth information gained from observation and its popularity in both literature and in clinical practice. Ultimately, the MET is ecologically valid

with high face validity as it is a standardised version of everyday tasks that may happen in a typical day [125]. Moreover, I specifically chose to focus on the MET as my group had made our own computer tablet version of it prior to my PhD, and I was awarded a fellowship to develop and validate it.

Currently the OxMET is available only for research and service-related projects and not for direct clinical practice. To be regulated as a medical device, robust diagnostic and algorithmic evidence is needed to justify the incorporation of a tool into clinical practice and I provide some of this evidence.

In part one, I examined the convergence, acceptability, and feasibility of the OxMET in comparison to the MET-Home in 48 stroke survivors and 50 healthy controls. The purpose of this project was to develop an evidence base for use of the OxMET and its relation to other METs and everyday behaviour. In this way, I have provided evidence for the veridicality of the OxMET, that is, the degree to which a test statistically relates to real life functioning [117]. Again, an assumption is made about the validity of the MET-Home as being a good marker of everyday behaviour for us to assume that associations with it are links to everyday behaviour. Ideally, covert observation of everyday behaviour in different situations and scenarios for a period of time would be a gold standard to judge everyday behaviours and compare to OxMET performance, but this is generally not possible. Furthermore, covert observation is not standardised in the same way, without influencing the participant about what can and can't be done to be able to compare to other participants. The MET-Home presents a way to standardise everyday behaviour in a comparable way across participants to remove variance of unexpected events which may attenuate correlations. Additionally, the requirement for this project to use an adaptable, standardised, non-computerised or virtual MET meant that the MET-Home was the only appropriate version published at the time of this project. In this way, it was the best option as a comparison task to establish convergence.

Following from evidence of convergent validity of the OxMET, it was necessary to evidence the predictive ability of the OxMET as a next step in generating validity evidence. Predictive validity in this case was defined as whether the OxMET

predicts outcomes which are temporally distant in test administration. This type of evidence is necessary in the Standards [21] to establish whether inferences about future outcomes can be made. Where the OxMET is intended as a cognitive screen for executive dysfunction for use at hospital bedside, the future outcomes refer primarily to recovery post-stroke or rather successful functioning after discharge. A way to quantify recovery and abilities is to look at functional outcomes. Functional outcomes in a hospital setting are relevant to both healthcare professionals and stroke survivors. On the one hand, clinicians want to understand if further rehab is necessary and which cognitive impairments may hinder rehabilitation and thus recovery. On the other hand, stroke survivors want to know what they can expect to be like in the immediate term to a few months after stroke. It is not the case that the OxMET can possibly predict life after stroke completely, after all it is an executive function screen, but it should contribute in predicting outcomes related to executive dysfunction.

The current study was divided into two parts; 1) an investigation of functional outcomes in a sub-acute (<2 months post-stroke) stroke sample in a rehabilitation setting; and 2) an investigation of predictive validity for activities of daily living chronically (~>6 months post-stroke) in a community setting. Overall the study had clear results: the OxMET, when assessed around 30 days post-stroke can predict activities of daily living over and above change in disability, stroke severity, age, education, and MoCA (where the MoCA was administered at 6-months follow up, and OxMET was administered at baseline). Indeed it was interesting that the complex executive functioning task that is the OxMET would predict overall ability of stroke survivors in terms of their impairments, participation, activity, and well-being. Other research has found such associations, revealing how simple tasks relate to higher level activities of daily living [1, 15, 120, 279].

The novel aspect of this project, is that such a brief and basic screen could aid in explaining an individuals current and future level of functioning. What was important in this paper was the associations between OxMET and clinically relevant measures. Tests should have good predictive validity, otherwise the diagnosis of a

problem that doesn't have evidence to affect functioning is of much less importance to healthcare professionals and stroke survivors. It is hard to perceive benefit of such a tool, and therefore hard to justify the use of the tool in stretched and overworked healthcare systems.

In summary, this paper generated further validity evidence for the use of the OxMET. The results revealed an independent contribution of OxMET to predicting later activities over and above MoCA or demographics. This is useful for healthcare professionals when understanding the impact being classified as impaired on OxMET could have on future recovery. One limitation is that I did not consider other factors that may be of relevance such as quality of life or mood etc., though these have reliably been shown to relate to ADL [280, 281]. Similarly, I did not consider other potential predictors (e.g., other domain-specific impairments) on other cognitive abilities (e.g., language or attention) and to what extent this contribution is unique to the OxMET.

3.6 Validity of neuropsychological tests and limitations in the presented works

3.6.1 Tests used in validity analyses

When attempting to provide validity evidence, one has to establish the validity of interpretations of test performance, rather than validity of the test itself [21]. When validating the OCS, I provided evidence that interpretations of performance in terms of impairment classifications were comparable, and thus interpretable, in the same way as the in-person OCS. For the OCS-Plus, I provided evidence that OCS-Plus performance was associated with standardised neuropsychological performance, meaning that flagging impairments could be interpreted in similar ways to other neuropsychological tests which are currently used and validated. In the final chapter, I provided evidence that the OxMET was comparable to the MET-Home, an in-person alternative to the OxMET, and interpretations could be made in the same way.

For each of these validation studies, the same weakness is common; there is no gold standard test. The term *gold standard* has different origins, but can refer to the best quality instrument or literal worth in gold, being the highest worth item historically [282]. For psychometrics, it is argued that there are few, if any, gold standard references at all [282, 283]. It is argued instead, researchers should refer to a test as 'extensively validated' rather than a misleading indication that any given psychological test is the best quality example of a measure [283], as all measures are imperfect and fallible measures of a given construct [284]. In neuropsychology a 'full neuropsychological evaluation' is considered the best available gold standard [45, 285]. A full neurosychological evaluation involves comprehensive neuropsychological assessment which can take several hours, across several occasions, and involve experts integrating information across multiple cognitive and psychological tests as well as wider clinical context and personal background [286]. A comprehensive neuropsychological investigation is likely not tolerable, nor feasible, nor appropriate for many stroke survivors in acute or hospital settings early following stroke [45, 287, 288]. The terms 'neuropsychological test' and a 'neuropsychological assessment' differ in that the first measures a given ability via a single or select number of tests and the other integrates all possible information which is interpreted by a specialist and may include testing [289].

I did not compare each presented test to a comprehensive neuropsychological assessment. There are several reasons for this. The first is practical, in that conducting a full assessment was not within budget for the studies and no expert clinical neuropsychologist was available to interpret cognitive abilities of all 602 participants I included in my thesis research (n=40 + n= 347 + n= 98 + n= 177). A further practical limitation is due to the cost of neuropsychological assessments: e.g. the Neuropsychological Assessment Battery [290] which costs $2,964.00 [291], or the full Behavioural Assessment of the Dysexecutive Syndrome [139] which costs £699.85 [292], or the Repeatable Battery for the Assessment of Neuropsychological Status [178], which costs £356.99 [293]. Indeed, the same logic of gold standard applies itself to neuropsychological assessment batteries; there is no 'gold standard' assessment

batteries to be included in the gold standard comprehensive neuropsychological assessments [45]. Harmonisations of neuropsychological assessments have been attempted in the past [294], but the brief 30 minute neuropsychological evaluation proposed was heavily language based and would therefore be inappropriate for stroke survivors with difficulties in communication. Instead, I used the tools available to me which included some, but not all, of the tests included in these assessments. To further justify my choice of fewer tests which are closely matched to our tests, please see my discussion on false positives for impairment found by increasing the amount of tests administered, even on healthy controls in empirical chapter 2 discussion.

3.6.2 Statistical methods used in validity analyses

Not only do the tests used in comparisons matter, but the analyses used are paramount. There are multiple alternative methods that are appropriate to provide evidence for construct validity, in addition to the ones used in this thesis. Two will be discussed here as they are relevant and methods of establishing construct validity [295]; 1) Multitrait-Multimethod Matrix (MTMM) and 2) latent structure modelling.

Multitrait-Multimethod Matrix (MTMM) [295], in essence, is a correlation matrix with the diagonal being the reliability coefficient of different methods of measuring the same or different traits. To illustrate MTMM use, Campbell and Fisk [295] used the example of measuring three traits with three different methods. Correlating these traits and measures allows one matrix to reveal both convergent and divergent validity evidence. The MTMM method can also distinguish the influence of the method of administration/test from the concept being measured. This method of analyses could not be used in my investigations due to limited occurrence of multiple methods assessing the same constructs. The MTMM is however, very restrictive in using only independent methods and completely independent traits or constructs to be compared [296]. Moreover, the method assumes that traits/constructs and methods are not or minimally correlated which is not always the case [297]. This fundamental assumption of the method casts doubt on the ability to interpret the matrix in light of discriminant validity [297].

Alternative methods of establishing validity include examining the existence of, and interrelations between, latent structures/factors representing cognitive domains. Confirmatory factor analysis (CFA) cannot only be used in the establishment of an internal structure of a neuropsychological test or test battery [298, 299], but also in the cognitive domains or constructs assessed in comparison to different tests from different batteries [300, 301]. One example of using CFA was by Carlozzi, Horner, Yang and Tilley [302] who examined the internal structure of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) [178] mentioned earlier (and from which I used the figure copying task in empirical chapter 2). The RBANS has 12 subtests, from which a structure is hypothesised of five-factors (immediate memory, visuospatial-constructional, language, attention, and delayed memory) [302]. The authors used neuropsychological assessment data from 175 male veterans from memory clinics in the United States of America [302]. The CFA on RBANS data did not replicate the five-factor structure, invalidating the initial hypothesised structure [302]. Instead, the data revealed a potential two-factor structure, with the first factor accounting for over 80% of variance [302]. In this case, the authors provided evidence of *invalidity* of the hypothesised structure similar to previous research in different samples [303, 304].

Similarly, Vellone et al. [305] conducted a CFA on version 3.0 of the Stroke Impact Scale (SIS), an eight sectioned self-report questionnaire measuring strength, hand function, activities of daily living (measured both on a basic and instrumental level), mobility, communication, emotion, memory and thinking, and participation/role function [71]. Prior to Vellone et al. [305], the fractionable nature of the SIS and whether it does measure the eight domains purported to be measured had not been evaluated. Using 392 moderate stroke survivors with an average MMSE [48] of 20.2 and mRS [227, 228] of 3.2, the eight-factor model did not fit the data, and provided evidence of invalidity of the test [305]. This means interpreting the test cannot rely on 8 distinguishable factors but another model of performance. Using an exploratory factor analysis, the authors reported four-factors of 'physical', 'cognitive', 'emotional', and 'social participation' domains. In order to provide

construct validity of the new four-factor model, the authors compared the factor scores of the four factors to the mRS, the MMSE, the Short-Form Health Survey [306], and the Hospital Anxiety and Depression Scale [307]. Vellone et al. [305] only found weak to moderate correlations between Short-Form Health Survey sections, but significant and large associations were found between the 'cognitive' factor and MMSE score, between the 'emotional' factor and Hospital Anxiety and Depression Scale, and between the 'physical' factor and Barthel Index [223]. In this study, the eight-factor structure was not found, but the authors provided validity evidence for the four-factor structure.

A critical issue with both presented studies using CFA, is the ability of authors to move the criteria of validity evidence. That is, when one does not find the hypothesised structure, a new structure is explored and assumed to the 'correct' structure and used in future (as exemplified in both presented studies [302, 305]). The nature of confirmatory factor analyses is to *confirm* an a priori hypothesised structure within data and to provide parameters of the model fit to given data [308]. This method of structural equation model is aligned with the understanding of validity in the Standards, that is, to have a pre-defined construct (in this case a factor structure) and provide validity evidence for this [21]. Where CFA is swapped for exploratory methods within the same data set, and then CFA is applied again to appropriate the concept of *a priori* structure, this is undesirable practice [309], and produces unfalsifiabile validity evidence. Unfalsifiable, due to the nature of always being able to 'validate' the data through model fit if the models are overfitted to the data and appear optimistic [310].

Latent factor modelling of any kind was not used in the presented works in this thesis. This is for several reasons. Foremost, it was not the aim of any works to examine the internal structures. However, by not validating internal structures, the interpretation of test performance is affected. This weakness is particularly pertinent regarding the OCS and OCS-Plus tests, which each have multiple subtests purporting to measure distinguishable cognitive domains as evidenced in convergent/divergent analyses.

3.6.3 Samples used in validity analyses.

For context, I will first describe typical stroke demographics before discussing how this relates (or not) to the samples used in my research and the implications this may have for interpreting/generalisting results. In the UK the median age of a stroke survivor is 70 or 76 for males and females respectively, with 80-85% experiencing an ischaemic stroke [311, 312]. Globally, the vast majority of stroke survivors (76.88%) are aged 66 and above, with 51.63% of stroke survivors being male. In comparison to latest UK census data [313], 81.70% of the UK are White, followed by broad categories of Asian (9.30%), Black (4%), Mixed (2.9%), and Other (2.10%). For older adults closer to average stroke age in the UK (>65yrs of age), the ethnic diversity is different; White = 93.60%, Asian = 3.8%, Black = 1.4% [314]. Using UK biobank data (n= 6,680), one study examined the risk of stroke by ethnicity and found that Asian participants had the highest incidence of stroke (274/100,000, compared to Black participants (234/100,000) [315] and White participants (203/100,000). Black participants, however, had a higher hazard ratio of stroke compared to white participants even when adjusting for age, sex, genetic risk factors, lifestyles, and comorbidities [315]. Taken together, a representative stroke sample for the UK, may include participants between 70-76 years of age on average, where the majority experience ischaemic strokes, the majority are White, but there may be an overrepresentation of asian people to account for increased risk compared to white people.

Attempts were made throughout the presented projects to provide a representative stroke sample. This aim was largely met. The ethnicity of the samples were reflective of the UK ethnicity demographics for older adults with 90-94% of participants identifying as white in most presented projects. However, incidence of ischaemic strokes of ~70-80% across projects was lower than the stroke population averages (80-85%) [311, 312], across the presented works. Subsequently, the rates of haemorrhage were higher. There is a bias in stroke research regarding ischaemic stroke survivors participating. Prominently, there are differences between

ischaemic and heamorrhagic stroke which preclude recruitment, such as survival rates. Approximately 40% of haemorrhagic stroke survivors die within one month of stroke [316] which means studies examining stroke survivors outside of this window will not reflect all stroke survivors but only those who survive.

All presented research includes stroke survivors who were typically at least 30 days or more post-stroke, with the majority date of testing being greater than 6-months after the stroke event. Given this bias towards later recovery in stroke, it is not expected to have an over representation of haemorrhagic stroke, which is a testament to the inclusive criteria for participation. Note, whilst there is a slight over representation, this may not be substantial. Often inclusion/exclusion criteria are too restrictive, even in trials examining haemorrhagic stroke, leading to lower recruitment of haemorrhagic stroke than expected based on population or hospital statistics [317]. Haemorrhagic stroke survivors tend to have more severe strokes than ischaemic stroke survivors. In a large sample of 35,491 stroke survivors in Denmark taken from a national registry, 10.10% were haemorrhagic stroke, and the presence of haemorrhagic stroke was related to higher stroke severity [318]. The risk of mortality between ischaemic and haemorrhagic stroke were insignificantly different >=3-months post-stroke, perhaps explaining why there are more haemorrhagic strokes included in my research than typical in the number of stroke types acutely [318].

Stroke severity has mixed evidence for its ability to independently predict cognitive impairment in stroke, with more research evidencing an association than not ([319, 320], cf. [321]). Yet separate research studies, and my own, have found no independent contribution of stroke severity, measured by NIHSS, to longer term cognitive impairments [9, 270, 271]. One reason for limited findings may be the stroke severity of samples used. In all cohorts presented in this thesis, the severity of stroke was moderate, there were limited but present instances of severe stroke. More severe strokes were not included (as opposed to consciously excluded) from my research [322]. Often, those with aphasia are excluded from stroke research due to the potential for incapacity to consent to take part where they cannot communicate [322]. Indeed, those with aphasia were not explicitly excluded from all research in

this thesis, however, those without capacity to consent to take part were excluded. Those who are more readily able to consent to research by appearance of the ability to communicate a decision are more likely to be approached by researchers and take part in research [322]. Incapacity to consent is statistically related to more severe strokes, increased mortality whilst admitted, higher disability levels at discharge, and cognitive impairment [323, 324]. It is possible that due to the lack of severe strokes in the samples used that impairment classifications and proportions were not reflective of the stroke population and precluded wider variance in analyses. This limits the generalisability and strength of associations found.

3.7 Integrating validity evidence: directions for future research

The process of generating validity evidence never ends [21]. Multiple systematic reviews exist to collate, and give the current state of evidence, for the use of the MoCA in detection of cognitive impairments [325, 326] and other cognitive screening tools [61]. For the Tele-OCS, future research should compare Tele-OCS to in-person OCS in a larger and more diverse sample, including healthy ageing adults. By doing this, normative data can be generated for the Tele-OCS and compared to in-person normative data to establish any shift in subtask cut offs due to modality change. Further, comparison between OCS versions in a more impaired stroke cohort, and different stroke cohorts (e.g., those closer to stroke event, those who are more impaired cognitively, or physically, or both etc) would increase variance and confidence in results found. Moreover, a larger sample would allow examination of the internal structure of the OCS, providing some evidence (or not) of the Tele-OCS structure.

Internal structure validity is a necessary next step for Tele-OCS, OCS, and OCS-Plus. Indeed, recent research has explored the internal structure of the OCS [327]. Using data from 1,973 stroke survivors and 498 healthy ageing adults across the UK and Italy, Iosa et al. [327] conducted principal component analysis (PCA)

on the OCS. Principal component analysis is similar to factor analysis in that it is used for data reduction and identifying variance between variables, however, it does not assume an underlying structure and therefore is not used in confirmatory ways for internal structure, only to describe variance in items [328]. Principal component analysis is often confused with factor analysis and words are incorrectly used interchangeably [329], which was the case in the PCA analysis of the OCS data. For instance, the authors applied a PCA to the OCS data in order to 'identify the main OCS domains to solve some scientific and clinical issues related to this useful and valid screening tool' (p3p1, Iosa et al., [327]), yet used descriptive methodology, not methods used to identify the latent structures within a test for the purposes of validity [329]. With the assumption instead that PCA was appropriate for their aims, they found three, five, or six component structures of the OCS. According to the OCS test structure [38], the OCS is made up of five cognitive domains (e.g., language, number, memory, praxis, and attention), yet the five-component model found was made up of 'language and arithmetic', 'orientation', 'memory', 'visuomotor control', 'spatial exploration', and 'executive functions'. The results do not provide evidence of validity for the internal structure of the OCS against the originally intended domains [38]. Of note regarding the study by Iosa et al. [327] is that the authors do not discuss the use of PCA for their study nor acknowledge the limitations of this statistical approach.

Alternatively, Bisogno et al. [330] appropriately used PCA on OCS and NIHSS [331] data from 158 first-ever stroke event stroke survivors in Italy, to 'reduce the number of variables and describe the variability of behavioural deficits' (p3p8, Bisogno et al., [330]). Note, this was not a validation study of the internal structure of the OCS, but rather to assess low-dimensionality of cognitive and behavioural performance in stroke survivors. When combining OCS and NIHSS items, the authors found a three component structure, with components described as 'language, calculation, memory, praxis', 'motor left, inattention, egocentric neglect left, sensory, overall performance', and 'motor right, sensory, egocentric neglect right' [330]. Again,

this component structure does not match that of the intended domains in the OCS [38], but this is possibly due to the inclusion of the NIHSS.

The results of both structural investigations of the OCS do not invalidate the OCS, as neither were designed or implemented in a way which can be considered providing evidence of the internal structure of the tests. The first study did not employ statistical methods appropriate for establishing (either in an exploratory way or confirmatory way) the internal structure of the OCS [327], and the second study used PCA in a descriptive way in combination with a stroke severity measure [330]. Without this validity evidence, the need to validate the internal structure still stands. Furthermore, as there are no studies examining the internal structure of the OCS-Plus, nor the OxMET, the same inferences cannot be made for these tests.

The OxMET does not make assumptions about being multi or unidimensional. The OxMET is designed as a measure of the effects of executive dysfunction on everyday behaviour captured in a computer tablet. Executive function is an umbrella term for multiple cognitive processes [109]. The original multiple errands test was based upon the Supervisory Attentional System (SAS) [332] which posits that where difficulty and novelty are experienced in everyday scenarios, behavioural schema relating to those typical situations must be updated inline with the novelty, and this is governed by SAS, which is a slow process allowing for problem solving [130, 134, 332]. In-depth discussion of this model is found elsewhere [332, 333]. In an attempt to develop tests using the SAS model which are more predictive of everyday behaviour, Shallice and Burgess created the Multiple Errands Test and Six Elements Test [106]. In particular, the MET is purported to measure the SAS abilities to identify goals, generate plans, forward think to establish cues to future behaviour, and be able to use the cues when needed [333]. It is not the case that particular MET metrics are aligned to particular SAS processes, but that the general approach of the MET is intended to capture the features of SAS [333]. It is not easily possible to establish an internal structure of the MET or OxMET as there is no reference for the internal structure to be validated against.

The OCS-Plus is structured primarily around memory tasks and measures of executive functioning and attention, among instep tasks of naming and orientation [50]. A confirmatory factor analysis should be conducted on healthy adults and stroke samples in order to maximise variance, with focus on diversity of cognitive impairment in stroke. A two-factor structure would be expected given the design of the OCS-Plus and its reflective measurement model [186], however, such analyses may reveal a diverse factor structure. To avoid identifying a model on data and then overfitting the found model on the same data, future research should retain a train and test dataset, with similar characteristics, to first identify the structure then confirm the structure.

For the OxMET, several avenues of research can be explored. The OxMET could be compared to other digital/virtual versions of the Multiple Errands Test, not only on a psychometric level but to evidence acceptability and feasibility of one version over another [270]. Establishing the wider link to digital versions expands the evidence base for use of the OxMET in clinical practice and may provide further evidence for constructs measured. Another avenue of research is to compare the OxMET to an extensive executive functioning battery to examine whether different metrics within the OxMET differentially associate with different aspects of executive functioning. Current interpretation of OxMET metrics is based on face value, for example, OxMET 'omissions' reflects how often participants forgot to, or simply did not, complete tasks. Note there is no definition of how this may relate to executive functions. In the original psychometric validation of the OxMET, comparison of the OxMET with OCS-Plus, the Zoo Map test [139], and Pillbox test [152] was made, but no extensive examination of executive function was conducted.

An important future research direction affecting all tests is to assess the feasibility of testing in real clinical scenarios. When licenses are issued for both OxMET and OCS-Plus, licensees are asked for what they will use the apps, on who they will use them, and on what bases persons are selected to take part in the apps, among other questions on their role and setting. Currently, license holders of OCS-Plus and OxMET are being issued a survey of actual use, versus intended

use, of the apps, and are assessed on their perceived usefulness of the apps based on the System usability scale [334]. Taken together, both the questions and the survey are attempting to get information on use of the apps. Evidence is needed for feasibility in practice, including barriers to testing, the acceptability of the tests by patients and assessors, and time taken to run and interpret tests. Moreover, direct comparison between our apps in practice and reference standards in practice is necessary with a focus on diagnostic accuracy, time, and interpretability.

A recent study by Murphy et al. [335] independently (from my research team and collaborators) examined the use of the OCS on an acute stroke unit for stroke survivors within 5 days of stroke (on average). The authors found that near 30% of their sample did not complete at least one OCS subtest, often the broken hearts task or trails tasks [335]. The authors further examined the impact of language impairments on OCS performance, given the OCS is adaptable to communication difficulties. They found that only 54% of stroke survivors who failed either the picture naming or semantics task (failing neither of which constitutes a diagnosis of aphasia) completed the other non-language subtasks, compared to 84% of 'non-aphasic' participants [335]. Inferring feasibility of the OCS from this data suggests the OCS is not as feasible for those with language problems than as anticipated by the design of the OCS [38]. Individuals with severe aphasia are precluded from completing most cognitive screening tools because they are not suitable, as discussed in this thesis, and this problem is not specific to OCS. The findings from Murphy et al. [335] have further implications for the use of the OCS remotely with those with aphasia, and this should be investigated explicitly in future research. Furthermore, Murphy et al. [335]'s findings in a single acute stroke unit should be extended to examine feasibility across multiple sites and collect data on reasons for inclusion and exclusions of participants (including formal aphasia diagnoses), which were not collected by Murphy et al. [335].

Conclusion

The aim of this thesis was to showcase empirical research providing validity evidence for different cognitive screening tools. The presented works each advance neuropsychological research and practice in similar ways. Generating a timely and essential standardisation of the Tele-OCS, provided clinicians with a way to assess stroke survivors in heavily restricted circumstances. Further, providing evidence of validity of the OCS-Plus in stroke, now allows greater confidence in use of the OCS-Plus in this key target population. OCS-Plus is already being used globally with over 800 licenses issued internationally and three language translations, evidencing the need for such a tool. The OxMET is a relatively new task, already used internationally with over 300 licenses and two international translations and adaptations. Evidence of OxMETs construct and predictive validity, therefore, also gives confidence to interpreting its output.

I have found that all measures examined in this thesis have multiple grounds for valid use in clinical practice and each have strengths and weaknesses in different settings and modalities. The overall recommendation of this thesis is to use any test, including the ones presented here, on the sample they were designed for, in the way they were designed to be used, and only when they were designed to be used (i.e., in acute or community settings etc). Importantly when all of those criteria are fulfilled, to interpret the test results in the intended way, and not to go beyond the evidence behind interpretation. None of the tests presented in this thesis can fully capture the depth of human cognition nor the rich idiosyncrasies of real everyday behavior. Health care professionals should always follow the latest guidance and be educated in best evidence-based screening use and trained in using the tools

they administer. When the latter is also fulfilled, stroke survivors stand the best chance of receiving the best screening and diagnostics.

4
Appendix A

All outcome metrics from main measures from our lab included in this thesis.

4.1 OCS

The following text is taken directly, with permission, from the supplementary materials of Demeyere et al's [38] (http://dx.doi.org/10.1037/pas0000082) paper, the beginning and end of the direct quote is signified by '- -' in the paragraph break. '- -'

– –

Language: Picture Naming

This is positioned as the first subtest to broadly assess level of expressive language, so helping to determine the administration of other tests (e.g., whether forced-choice procedures will be required). In addition, since picture naming incorporates recognition of the visual structure of the object, access to semantic/ conceptual knowledge about the object, word finding and articulation, the task taps several different processes. The items are chosen to cover a range of difficulty and to be suggestive of any problems between different categories (see Testbooklet). The patient is separately presented with 4 pictures to name (hippopotamus, water

melon, filing cabinet and pear). Each picture is centred in the middle of the page to reduce confounds by spatial neglect. Participants score 1 for each correct answer (maximum score = 4). Melon and chest of drawers are allowed alternative responses (to water melon and filing cabinet) and should also score 1. Self-correction is permitted and the final answer is taken as the participant's response. The score contributes to the overall assessment of language.

Language: Spoken word comprehension

This task is administered immediately after picture naming, to assess comprehension levels in a patient. In patients with aphasia, the task allows assessing of basic receptive abilities and the ability to follow simple instructions (pointing to a picture). The patient is presented with 4 pictures simultaneously, arranged in a vertical orientation to minimise effects of spatial neglect (see the Test booklet). The patient is asked to point to the fruit, then the animal and then the tool. The participant scores 1 for each correct answer (maximum score = 3). Self-correction is permitted and the final answer is taken as the participant's response. The score contributes to the overall assessment of language.

Memory: Orientation

After having assessed levels of expressive and receptive language ability, the patient is briefly screened for orientation in time and space. Patients are asked about which city they are in , the part of the day, the month and the year. First an open question is asked with free response from the participant. If the participant cannot respond (e.g. due to aphasia), makes an error or does not know, multiple choice options are presented. When the multiple choice options are given, the question is read out to the participant and four possible alternative choices are also read out and pointed to by the examiner to direct the participant's attention to the items in succession. The participant can then point to one of the alternatives to make a response. In order to allow a level playing field for patients with aphasia, there are no penalties for responding to these questions using the multiple choice pointing

options. The participant scores 1 for each correct answer (maximum score = 4). The score contributes to the overall assessment of Memory.

Visual Fields

Given that hemianopia can significantly affect functional recovery (Ramrattan et al., 2001), the OCS includes a simple confrontation test to assess the patient's visual fields. The examiner faces the participant and raises both hands, first to the participant's upper visual fields, and then to the participant's lower visual fields. In each of these locations, the examiner is asked to wiggle the fingers of either the left or the right hand, while checking that the participant maintains central fixation on the examiner's nose. The participant is asked to indicate by pointing which hand they saw move. One point is scored for each quadrant, resulting in scores out of 2 for the left and right visual hemifield. Any deficit found here could inform further test performance, for example by indicating that care should now be taken to ensure that all subsequent tests are performed in the intact visual field.

Language: Sentence Reading

A single written sentence is presented, centrally aligned on the page and the patient is asked to read the sentence out loud. There are 15 words arranged in 4 rows (3 with 4 words each, one with three). The sentence contains 4 critical irregular words (islands, quay, colonel, and yacht in version A; favourite, mauve, brooch, and trough in version B). As well as determining whether the patient is able to read out loud in general, measures of more specialised dyslexias can be detected for further assessment. For example, regularisations of the 4 irregular words would suggest surface dyslexia. The reading task may also indicate neglect if the patient fails to read words on a specific side of space or if an error is made on one side of a word (e.g., reading islands as hands). After the participant has read or attempted to read the sentence, the examiner then reads the complete correct sentence to the patient. This is to ensure the correct sentence has been heard by all patients as the sentence is subsequently used for a surprise verbal memory task, later in

the screen. Patients with expressive aphasia have the sentence read to them twice in order to allow a level playing field for the later recall/recognition test. The participant scores 1 for each correctly read word (maximum score = 15). The score contributes to the overall assessment of Language.

Number tasks

There are 2 sections. In a number writing task, the patient is presented with a piece of lined paper and is asked to write down, in figures, the number they hear. The numbers are: "708", "15,200" and "400". These items are chosen because 708 and 15,200 require correct syntactic parsing (e.g., to prevent writing 7008) while writing 400 tests if a syntactic error (writing 7008 or 15000200) reflects a perseverative response (in which case 400 might be written as 4000). The numbers may be repeated once if necessary. The maximum score is 3 (1 point per number). The second task requires simple calculations - 2 additions and 2 subtractions. The sums are presented visually in a large font, centred on the page for free responses – patients are also allowed to write down their answers. If a free response cannot be given (e.g. due to expressive dysphasia), the patient is asked to select (by pointing) which one of 4 possible answers is correct. In order to allow a level playing field for patients with aphasia, there are no penalties for responding to these questions using the multiple choice pointing options. The participant scores 1 for each correct answer (maximum score = 4). The scores contribute to the overall assessment of Number.

Visual attention: Broken Hearts (Hearts Cancellation)

The task is to cancel out all complete hearts while not crossing out hearts that are broken (i.e. with a gap on the left or right side). The test assesses egocentric neglect by comparing the number of complete hearts selected on the left versus the right side of the page. In addition, allocentric neglect is measured when patients consistently make false positive responses to distractors with a gap on one side (see Bickerton et al., 2011; Chechlacz et al., 2012). The patient is first presented with a practice page containing an example of the task and 7 practice hearts, aligned centrally on the

page to minimise effects of neglect at this stage. Feedback is given and the test is started when a patient makes a reliable selection of the target hearts. If the patient makes no selection, or selects all of the hearts, a second set of practice hearts is given. When the practice is completed, a page with 150 semi-randomly positioned hearts is placed in front of the patient and the patient is given a maximum of 3 minutes to complete the task. The total number of complete hearts cancelled is recorded. An asymmetry score is also computed by subtracting the number of hearts cancelled on the left of the page vs. the number cancelled on the right.

Praxis: Imitating meaningless gestures

This test requires that a patient mirror imitates, with their better hand, a series of meaningless hand and finger actions made by the examiner. There is very little language demand and, as the better hand is used, the task does not rely on motor strength. Similar imitation tasks have previously been shown to begood predictors of apraxia (see Humphreys et al., 2012; also Bickerton et al., 2012). There is a short practice requiring mirroring the lifting of a hand, after which the patient is asked to copy two sequences of hand positions relative to the head. If the patient makes an error or is imprecise the gesture sequence is repeated once. This is followed by the imitation of two finger positions which can be repeated if the initial response is wrong. The task is scored out of 12 points in the following way. For each gesture sequence three points are given if the gestures are correct and precise after 1 presentation, 2 points if they are correct and precise after the second presentation, 1 point if there is one error after the second presentation, and 0 points are given if, after the second presentation, there is more than one error, no response or perseveration from a previous item.

Memory: verbal memory

A surprise verbal memory task is given, asking the participant to recall the sentence that was read/spoken earlier on in the screen. Only the recall of the irregular words (target words) is scored. If the patient is unable to recall all four target words,

multiple choice assessments are given for each missed or incorrect target word. Four vertically aligned options are shown and read from a page in the Test Booklet. These options consist of the target word, a semantically similar distractor and two unrelated words. The examiner points to each word in turn and reads it aloud "Can you point to the word you remember from the reading test or perhaps the one that looks more familiar to you?". If the patient had been able to recall any words from the sentence correctly, their recognition of those words is not assessed. This makes up the verbal memory part of the OCS. A further 4 multiple-choice questions are then given to address non-verbal, episodic memory through task recall. For the episodic memory test the patient is asked to point to which of 4 vertically aligned items had been encountered earlier. Four sets of questions are given.

Controlled Attention (Set Shifting): OCS Trails task

The final sub-test aims to measure executive function. This uses a trail-making task designed to minimize language demands and also to extract-out the effects of neglect on performance. The task requires patients to draw lines between simple shapes. There are 3 test components. There are two baseline tasks ((i) connecting circles in decreasing order of size, in the presence of triangle distractors, and (ii) connecting triangles in decreasing order of size, in the presence of circle distractors). The baselines are compared with a set shifting condition in which participants draw a trail alternating between circles and triangles, each going down in order of size (largest triangle to largest circle to next largest triangle etc.). The items are positioned semi-randomly, so that a correct trail can be drawn without going through any of the other shapes, and the items appear in a central section on the page. Performance is timed. The effect of set shifting is assessed by subtracting performance in the set shifting condition from that in the baselines. This subtraction extracts out effects of neglect, as items can be missed on one side of the page in both the experimental and baseline conditions, and motor slowness (present in the baseline as well as the set shifting condition).. For scoring, one point is given for each correct connection (if an error is made at some point, but subsequent performance

is correct, the correct connections are acknowledged). Prior to each proper run in one condition, patients are given a visual demonstration of the task to ensure there is task comprehension and that effects of poor verbal comprehension are minimized.

– –

4.2 OxMET

Task metric descriptions are presented in Table 7.8.

4.3 OCS-Plus

The OCS-Plus tasks include OCS tasks and new ones taken from both the Birmingham Cognitive Screen [336] and novel tasks. See Table4.1 for a description of metrics and tasks in the Oxford Cognitive Screen - Plus (OCS-Plus).

Table 4.1: Brief subtask descriptions and order of administration within OCS-Plus

Task order in OCS-Plus	Task name	Brief description	Scoring/metric
1	Picture Naming	The test includes visual object recognition and access to semantic/conceptual knowledge about the objects, word finding, and articulation. The items are chosen to cover a range of difficulty and to be suggestive of any problems between different categories. There are four low frequency items to name. Performance is relatable to language as well as reported under the memory domain (word finding). These items also form part of the incidental Episodic memory test further in the OCS-Plus	Four items are scored for correct response giving a score range from 0 – 4
2	Semantics	This Semantics test assesses both specific object and semantic category knowledge. This uses multiple choice matching within semantic categories of exemplar pictures to names. Performance is relatable to language as well as memory domain. These items also form part of the incidental Episodic memory test further in the OCS-Plus	Four items are scored for correct response giving a score range from 0 – 4
3	Orientation	Orientation in time and space, related to long term Memory. The participant is asked which year, month and date it is, and is then also asked whether they can name the current prime minister.	Four questions are scored for correct response giving a score range from 0 – 4

4	Word Memory Encoding	In the Word Memory test, the participant is given a list of five words to learn (bicycle, mist, wardrobe, teacher, and rectangle). The participant is asked to recall the items immediately after presentation, then regardless of performance the participant is presented with the word list again and asked for immediate recall a final time. This forms immediate verbal recall over two stages of Encoding.	Five words are scored for correct response giving a score range from 0 – 5, encoding is scored separately for first immediate recall and second immediate recall.
5.	Trails	The Trails test has three components. There are two baseline tests: (i) connecting circles in decreasing order of size, in the presence of square distractors, and (ii) connecting squares in increasing order of size, in the presence of circle distractors. The baselines are compared with a shape switching condition in which participants draw a trail alternating between circles and squares, with circles going in descending order of size and squares in ascending order of size. The items are positioned semi-randomly, so that a correct trail can be drawn without going through any of the other shapes, and the items appear in a central section on the page. Performance is timed.	For baseline squares and circles test trials are both scored for total correct connections out of seven. Scoring is not all of nothing, if someone continues with the rule following an error, one error is recorded and thus one less accuracy point.

		The effect of shape switching is assessed by taking performance on the switching task over performance on both baseline tasks summed.	Processing speed was calculated as the sum time taken on both versions of the baseline tasks (circles or squares) divided by proportional accuracy. This took into account how few or many targets were connected, and the time taken to do so. Removing impossibly quick processing speeds from those who only connected a few stimuli.
6	Verbal Recall	The participant is asked again to reproduce the list of words they learned in the Word Memory test. If the participant is unable to recall all words correctly, multiple choice assessments are given for each missed or incorrectly recalled target word to test recognition. For this, four vertically distributed options are shown and read to the participant and they are asked to make a forced choice response.	Five words are scored for correct free recall giving a score range from 0 – 5
7	Episodic Recall	A further four multiple choice questions address the participant's visual and Episodic memory through recognition of previously presented items, pictures, and words.	Four items are scored for correct response giving a score range from 0 – 4

8	Figure Copy/Recall	The participant is asked to copy a composite image of geometric shapes, with the image to be copied initially present at the top of the page for the duration of the copying. Subsequently, and immediately after, the participant is asked to daw the same composite figure once more, this time from memory.	Scores are given per specified element (20) drawn in terms of: presence, position, and overall accuracy. Each element gets a score out of 3 and therefore the total score possible score is 60 for both Figure Copy and Figure Recall
9.	Rule Finding	The participant is presented with three columns of alternating geometric shapes (squares-triangles-squares), rows of alternating luminosity (dark-light), and a red dot that moves around the pattern following certain rules. The participant has to try and pick up the rules to predict where the dot is going to go next based on previous moves (the most recent preceding position is highlighted on the display). The rules will change throughout the test without giving any notice and it is the participant's task to pick up on the change and work out the new rule. There are five rules in this test.	Two scores possible. If at least two consecutive guesses are correct within the same rule then the rule is scored as learned. There are five rules. Total number of correct guesses, excluding those immediately after a rule change, give a total accuracy score out of 46.

10	Cancellation	Search task including semantic category items, assessing organized search. The participant is asked to select drawings of fruit amongst drawings of common fruit and vegetables immediately followed with an invisible version of the same cancellation display. In the version administered first (feedback version) the selected drawings are framed, and this visual feedback stays on the screen for the duration of the test. In the version administered second (no feedback version) the visual feedback is only visible for the duration of the pen tap and it disappears afterwards; this requires the participants to remember selected items and try not to revisit Cancellations	Total number of correct fruit selections in both the visible and invisible feedback conditions provide a score range of 0 – 30 False positives are scored as 1 error point per incorrect item selected in the visible cancellations task Correct revisits are scored as 1 error point per time an item is reselected in the invisible feedback conditions

5
Appendix B

Appendix items for chapter 1

5.0.1 Altering of participant data for items 9 and 10

Before analysis we investigated whether the data were correct, as such we examined answers to item 9 through to 10.b. We found 14 respondents whose responses to items 9.a-b did not sum to 100%, which they were advised to check, of these 11 did not sum to 100% because they instead summed to number of tests administered (e.g., they administered 10 tests, and said seven were face to face, two were via phone and one was on video call). On item 10 we found 31 respondents whose responses to items 10.a-b did not sum to 100%, which they were advised to check, of these 19 did not sum to 100% because they instead summed to number of tests administered (e.g., they administered 10 tests, and said seven were face to face, two were via phone and one was on video call). Following our preregistration, we examined the remaining participants whose responses for items 9 and 10 did not sum to 100% or the number of tests administered, to check if there were obvious ways to rectify their data, we decided to exclude these participants from the comparison analyses between pre-March 2020 and post-March 2020, but to include them in other analysis. We transparently report sample size per descriptive or analysis conducted.

5.0.2 Tables for COVID-19 survey paper

Table 5.1: Coding scheme used for scoring the frequency of themes expressed in the qualitative analysis of survey data (items 12-15, and 17)

Code	Theme	Item 13 (%)	Item 14 (%)	Item 15 (%)
2.0	Patients/participants can't be tested remotely due to lack of technology/know how/anxiety		28.87	
3.0	Lack of standardised validated measures (inability to access)		27.84	
3.1	Lack of standardised validated measures (few exist for purpose)		26.80	
4.0	Remote testing would not be valid due to other concerns		18.56	
5.0	No need for remote assessment			
6.0	Standardised remote assessment is not yet available		17.53	
7.0	Standardised remote assessment available, but just not used yet			
8.0	Unsure which tests to use remotely		14.43	
9.0	Concerns re reliability / stability of connection during assessment / technological issues		14.43	
10.0	Cannot compare remote tests to f2f tests			
11.0	Not familiar with the tech to conduct remote assessments			
12.0	Concerns regarding security		14.43	
13.0	Unaware remote assessments could be used			
14.0	Struggle with engagement		12.37	
15.0	Unable to remotely assess individuals with types of cognitive impairment affecting practical ability to take part		9.28	
16.0	Participant has other issue that may make remote assessment impractical such as being blind or deaf		9.28	
17.0	Cannot control distractions/environment		9.28	
18.0	Cannot do functional assessment	4.12	7.22	
19.0	Tests are not doable remotely		6.19	
20.0	Hard to establish rapport	2.06	6.19	
21.0	Harder to stop others helping the participant	2.06	5.15	
22.0	No normative data for remote assessments	43.30	5.15	
23.0	Lack of technology		4.12	
24.0	Cannot control the participant recording, finding answers, or cheating another way	21.65	4.12	
25.0	Staff/examiners not confident/knowledgeable on remote testing	4.12	4.12	
26.0	Difficulty picking up behavioural issues / interpreting behaviour/subtleties of behaviour		4.12	
27.0	Access to suitable place to test (quiet office etc)		3.09	
28.0	Difficult to manage distress		3.09	
29.0	Less personal		3.09	
30.0	Costs associated		2.06	

Table 5.1: Coding scheme used for scoring the frequency of themes expressed in the qualitative analysis of survey data (items 12-15, and 17) *(continued)*

Code	Theme	Item 13 (%)	Item 14 (%)	Item 15 (%)
31.0	Communication barriers over internet		2.06	
32.0	Cannot easily help with issues at other end (technical or physical)		2.06	
33.0	Can take longer/more complex to assess when having to set up and administer in a new way		1.03	
34.0	Test choice limited		1.03	
35.0	Requires others to help		1.03	
36.0	Does not work in inpatient facility		1.03	
101.0	Increased accessibility on the side of the participants attending appointments			
102.0	Safety			
103.0	It doesn't halt some existing care should lockdown occur again			
104.0	Time / economic /environment savings		1.03	
105.0	Easier to assess certain groups (e.g., non ambulatory or easily distracted by new things)			
106.0	Participant may prefer it/ be more comfortable with remote assessment			
107.0	Able to see function / see more at home than in clinic	4.12		
108.0	Less travel (rather than less time travelling)			
109.0	Flexibility of appointments	2.06		
110.0	Electronic information storage	2.06		
111.0	Less travel related effects on the individual (such as fatigue, anxiety etc)	43.30		
112.0	No PPE which could affect administration of tests			
113.0	Greater availability of professionals (e.g., an individual could access a neuropsychologist remotely but not in person in normal times)	21.65		
114.0	Able to ask friends/relatives extra questions that might not be possible if they don't all attend appointments	4.12		
115.0	Increased frequency of care			
116.0	Access a greater number of people to assess			
117.0	Better than nothing at all			
118.0	Able to follow up in community			
200.0	Standardised, validated, and normed tool creation			28.87
202.0	Access to remote administration software			13.40
203.0	Access to technology such as webcams			13.40
204.0	Technical support for the participant			13.40
205.0	Physical support for the participant			11.34
206.0	Increased time to set up / administer tests			7.22
207.0	Access to suitable environment to administer tests/ assess someone			6.19
208.0	Use of additional tests before or after remote assessment to get wider picture			5.15

Table 5.1: Coding scheme used for scoring the frequency of themes expressed in the qualitative analysis of survey data (items 12-15, and 17) *(continued)*

Code	Theme	Item 13 (%)	Item 14 (%)	Item 15 (%)
209.0	Low cost of validated/normed remote assessment tools			4.12
210.0	Access for participants to resources (i.e., they are available)			4.12
211.0	Contact prior to assessment to explain what will happen/how			4.12
212.0	Ensuring the participant is not helped			3.09
213.0	Training in use of remote assessments			3.09
214.0	Internet /tech access for the participant	4.12		2.06
215.0	Easy to administer tests			2.06
216.0	Availability of remote assessments	2.06		2.06
217.0	Seeing the participant face to face at least once	2.06		2.06
218.0	Guides on how to take part in remote assessment for participants	43.30		1.03
219.0	Better communication by the assessor			1.03
220.0	Multimodal remote assessment (video and photo or phone and photo etc)	21.65		1.03
221.0	Test administration that doesn't violate copyright	4.12		1.03

Table 5.2: Cognitive assessments used remotely prior to March 2020 and tests used since March 2020

Cognitive test	Administer prior to March 2020 (%)	Administered after March 2020 (%)
Montreal Cognitive Assessment	16.35	23.08
Delis-Kaplan Executive Function System	12.50	0.00
Oxford Cognitive Screen	10.58	7.69
Addenbrooke's Cognitive Examination	9.62	7.69
Wechsler Adult Intelligence Scale	8.65	15.38
Wechsler Memory Scale	6.73	0.00
California Verbal Learning Test	4.81	0.00
Mini-Mental State Examination	4.81	0.00
Test of Premorbid Functioning	4.81	7.69
Repeatable battery for the Assessment of Neuropsychological Status	3.85	0.00
Rey Auditory Verbal Learning Test	2.88	15.38
Wechsler Intelligence Scale for Children	2.88	7.69
Abbreviated Mental Test	1.92	0.00
Boston Naming Test	1.92	0.00
Doors and People	1.92	0.00
Hayling test	1.92	0.00
NEPSY-II	1.92	0.00
Non-standardised test	1.92	0.00
Wechsler Abbreviated Scale of Intelligence	1.92	0.00
Wechsler Individual Achievement Test	1.92	0.00

Table 5.2: Cognitive assessments used remotely prior to March 2020 and tests used since March 2020 *(continued)*

Cognitive test	Administer prior to March 2020 (%)	Administered after March 2020 (%)
Advanced Level Certificate in English (ALCE – reading)	0.96	0.00
Behaviour Rating Inventory of Executive Function	0.96	0.00
Brixton test	0.96	0.00
Cattell's Fluid Intelligence Test	0.96	0.00
Clinical Evaluation of Language Fundamentals	0.96	0.00
Cognitive Linguistic Quick Test	0.96	0.00
Colour Word Interference	0.96	0.00
Controlled Oral Word Association Test	0.96	7.69
Developmental Profile-3	0.96	0.00
Digit span	0.96	0.00
Graded Naming Test	0.96	0.00
Hospital Anxiety and Depression Scale	0.96	0.00
Language for cognition screening assessment	0.96	0.00
Neuropsychological Assessment Battery	0.96	0.00
Orientation log	0.96	0.00
Orpington Prognostic Scale	0.96	15.38
QuickScreen Dyslexia Test	0.96	0.00
Rey–Osterrieth Complex Figure Test	0.96	0.00
Rookwood Driving Assessment	0.96	0.00
Stanford–Binet Intelligence Scales-5	0.96	0.00
Telephone Interview of Cognitive Status	0.96	0.00
Test of Word Reading Efficiency	0.96	0.00
The Alzheimer's Disease Assessment Scale-Cognitive Subscale	0.96	0.00
The Autism Diagnostic Interview-Revised	0.96	7.69
The Gray Oral Reading Test	0.96	0.00
Verbal fluency	0.96	0.00
Visual Object and Space Perception Battery	0.96	0.00
Wechsler Test of Adult Reading	0.96	0.00
Wide Range Assessment of Memory and Learning	0.96	0.00

6
Appendix C

Appendix items for chapter 2

Table 6.1: Summary of construct and format matched validation comparisons for the OCS-Plus

OCS-Plus subtask	OCS-Plus metric(s)	Validation test(s)	Validation metric(s)	Validation metric(s) cut offs
Picture Naming	Number correct (0-4)	OCS picture naming	Number correct (0-4)	<3(5th centile)(3)
		MOCA picture naming	Number correct (0-3)	-
Semantics	Number correct (0-3)	OCS semantics	Number correct (0-3)	<3(5th centile)(3)
Orientation	Number correct (0-4)	OCS orientation	Number correct (0-4)	<4(5th centile)(3)
		MOCA orientation	Number correct (0-6)	-
Word Memory	Encoding 1 Number correct (0-5)	MOCA word recall	Number correct (0-5)	-
	Encoding 2 Number correct (0-5)	MOCA word recall	Number correct (0-5)	-
Trails	Trail A accuracy (averaged correct connections in trail A circles and squares) (0-7)	OCS Trails	Trail A (averaged correct connections in trail A circles and triangles) (0-6)	-
		CLTQ symbol trails	Number correct connections in Trail A (averaged from shape and size trails) (0-5)	-
	Trail A time (averaged completion time in trail A circles and squares) (0-inf)	OCS Trails	Trail A (averaged completion time in trail A circles and triangles) (0-inf)	-

		CLTQ symbol trails	Completion time in Trail A (averaged from shape and size trails) (0-inf)	-
	Trail B accuracy (0-14)	OCS Trails	Trail B correct connections (0-13)	-
		CLTQ symbol trails	Number correct connections in Trail B (0-10)	aged >=70 = score <6; aged <70 = score <9)(29)
	Trail B time (0-inf)	OCS Trails	Trail B completion time (0-inf)	-
	Executive score (0-100)	OCS Trails	Executive score (-13 - 12)	>4(5th centile)(3)
Word Recall	Number correctly recalled and number recalled + recognised from options (0-5)	OCS sentence recall and recognition	Number correct (0-4)	<3(5th centile)(3)
		MOCA word recall	Number correct (0-5)	-
Episodic Memory	Number correct (0-4)	OCS episodic recall	Number correct (0-4)	<3(5th centile)(3)
Figure Copy	Copy total score (0-60)	RBANS Figure	Copy total score (0-20)	<14 (sample M-2SDs)(30)
	Recall total score (0-60)	RBANS Figure	Recall total score (0-20)	<5 (sample M-2SDs)(30)

Rule Finding	Number correct (0-46)	Brixton spatial anticipation task	Number correct (0-46)	2SDs above the mean per age category (aged <60 =>29.9 errors; aged 60-69 =>31.4 errors; aged 70-79 =>34.5 errors; and >80 = >38.7 errors)(27)
	Number rules learned (0-5)		Number rules learned (0-9)	-
	Completion time (0-inf)		Completion time (0-inf)	-
Cancellation	Number correct (0-30)	BIT star cancellation	Number correct (0-54)	<51 (5th centile)(25)
	Number false positives (0-30)		Number false positives (0-79)	-
Invisible Cancellation	Number correct (0-30)	BIT star cancellation	Number correct (0-54)	<51 (5th centile)(25)
	Number correct revisits (0-inf)	Corsi block test	Number correct (0-7)	<5 (5th centile)
Domains impaired	Number of tests impaired on the OCS-Plus	MoCA	Total score on the MoCA	Either <23(31) or <26(12),

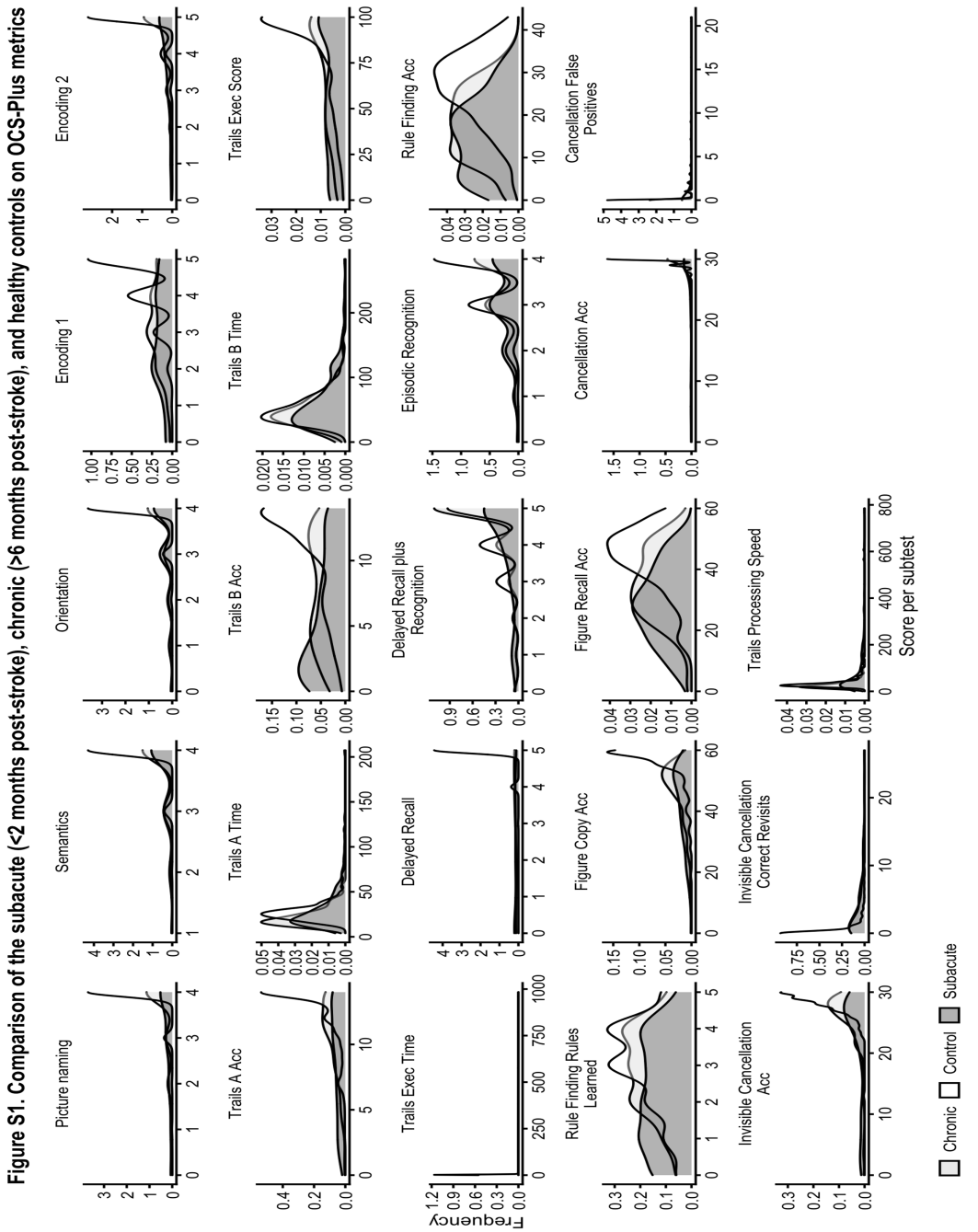


Figure 6.1: Comparison of subacute (<3 months post-stroke), chronic (>6 months post-stroke), and healthy controls on OCS-Plus metrics. Figure available under CC-by 4.0 license https://osf.io/796m3/

7
Appendix D

Appendix items for chapter 3

Reasons for non-participation in the current study: Healthy control sample: lives too far away = 47.83%; not a good time to participate for other reasons than illness = 17.39%; not interested = 30.43%; and passed away = 4.35%. Stroke sample: feels unable to complete a portion of the task: doesn't want to be seen at home = 2.78; feels unable to complete the MET-Home portion of the task: no where to carry out errands in residence = 2.78; feels unable to complete the MET-Home portion of the task: reports mobility is too poor = 8.33; no answer to contact = 44.44; not a good time to participate for other reasons than illness = 11.1; not interested = 11.1; passed away = 2.7; and too unwell to participate currently = 16.67%.

7.1 New normative data and concurrent validity for the MET-Home revised in english speaking healthy adults and stroke survivors

7.1.1 Abstract

Objective The Multiple Errands Test - Home (MET-Home) is a home-based standardised assessment of executive dysfunction. I developed a revised version

(MET-Home revised) to combat poor acceptability of two items. I aimed to psychometrically validate the new version. Crucially, I aimed to provide normative data for clinician interpretation. **Method** I combined data from two previous studies using the MET-Home in the USA [126] and UK (previous section), and compared the original MET-Home to the revised version. I accounted for covariates of age, education, mobility, disability level, stroke severity and days since stroke in comparisons. I assessed internal reliability. For convergent validity, I correlated MET-Home with global cognitive functioning (Montreal Cognitive Assessment), basic functioning (Barthel Index), and instrumental activities of daily living (Nottingham Extended Activities of Daily Living scale). **Results** Data from neurologically healthy participants (n=78) and survivors of stroke (n=66) were analysed. MET-Home versions were not statistically different in accuracy, omissions, or partial completions (all $p>.05$). MET-Home reliability was high ($a=.80$). Known group discrimination was good and specificity of MET-Home accuracy was high (90% specificity), but sensitivity was low (30% sensitivity). The MET-Home was highly related to global cognitive functioning ($r=.56$, $p<.001$), instrumental activities of daily living ($r=.46$, $p<.001$) and basic functioning ($r=.35$, $p<.001$). **Conclusion** I encourage use of the MET-Home revised for greater participant acceptability over use of the original MET-Home. The MET-Home is a well validated and newly normed tool useful for clinical investigation of executive dysfunction, but mainly for identification of no impairments.

7.1.2 Introduction

The original investigation of the MET-Home included 46 age, education, and gender matched community dwelling chronic (M= 1.5yrs post-stroke) survivors of stroke and neurologically healthy adults [126]. Burns et al. [126] screened participants for dementia [337], functional independence [223], depression [102], and for stroke survivors, their severity of stroke [331]. Their project established the internal consistency and inter-rater reliability of the MET-Home, which was found to be high (all 14 items $a=.73$, and inter-rater reliability was $> .88$). They established

convergent validity of the MET-Home through comparison to an ecologically valid performance based assessment of executive function [338] and a measure of processing speed [339]. However, significant convergent associations were not found between the MET-Home performance metrics and the Montreal Cognitive Assessment (MoCA) [37] as a measure of global cognitive impairment, structured and abstract executive functioning assessments (i.e., the Delis–Kaplan Executive Function System; Delis et al., [177]), and self-reported executive and emotional impairments in daily life (i.e., the Dysexecutive Questionnaire from the Behavioural Assessment of Dysexecutive Syndrome; Wilson et al., [139]). The non-convergent findings, albeit underpowered, suggest that the MET-Home may measure objective performance based executive dysfunction that affects everyday life, but it may not capture basic executive processes which are captured in structured and abstract neuropsychological testing. The authors suggested that as real life tasks are complex and multi-dimensional [340] and that structured neuropsychological tests individually do not capture this, the MET-Home may capture the reality of executive dysfunction well [126].

In a previous section I presented work comparing the MET-Home to the computer tablet based Oxford Digital Multiple Errands Test (OxMET) [130]. I collected data from 98 participants (48 stroke survivors [M=1.38years post-stroke] and 50 healthy ageing controls) on performance on both tasks, feasibility of both tasks in terms of predicting completion and completion rates, and some qualitative feedback from participants on acceptability. The MET-Home converged with the OxMET, and was found to be feasible. Participants accepted the MET-Home but commented on how challenging it was and that some tasks should be replaced. The study reported performance data on the original and revised version of the MET-Home, but they did not report on the revision process and whether this changed the validity of the task.

In the current study, I document the revision process, add to data collected previously, compare MET-Home versions, provide some further reliability and validity evidence, and present preliminary normative data. Furthermore, I wanted to establish the relationship between the MET-Home and MoCA in a larger sample to avoid low power encountered in the original investigation, as well as activities of

daily living and functioning. It was previously noted that the Barthel Index as a functional independence measure had only been used as a screen for MET studies, not as an outcome measure. As such, as a departure from Burns et al. [126], I used the Barthel Index as an outcome to establish any relationship between MET-Home and functional independence. My overarching aim was to provide justification to clinicians for which MET-Home version to use, to support the MET-Home's use cross-culturally (albeit still in a western culture), and to give clinicians the first benchmark for inferring impairment on the MET-Home.

7.1.3 Methods

How the sample size was determined, any data exclusions (if any), all manipulations, and all measures in the study are reported (Solution for transparency by Simmons et al. [171]). The COSMIN guideline for studies on measurement properties [217] and the STROBE cohort checklist [188] were adhered to in this manuscript. Ethical approval was granted by the Medical Sciences Interdivisional Research Ethics Committee (REF: R73921/RE001) of the University of Oxford. Written informed consent was provided by all participants.

Study design

A prospective cross-sectional design was utilized in four short phases: 1) initial deployment of the original MET-Home, 2) revision of MET-Home, 3) deployment of MET-Home revised, and 4) continuation of data collection for normative analysis. Phases 1-3 were reported in a previous section, while phase 4 is reported in this document.

Participants

Secondary MET-Home data (N=144) was collated from the original MET-Home study (N=46, n=23 stroke, n=23 healthy ageing) (Burns et al., [126]) and a subsequent comparison to the OxMET (N=98, n=55 healthy ageing, n=43 stroke; Webb & Demeyere [270]). Additional participants (n=2 stroke, n=5 healthy

controls) who participated in a student project since the conclusion of the OxMET comparison study were also included [270].

Measures

Briefly described, the measures relevant to the study objectives are: MET-Home [126], Montreal Cognitive Assessment (MoCA), Barthel Index (out of 20), Nottingham Extended Activities of Daily Living (NEADL), stroke severity as measured by the National Institute of Health Stroke Scale (NIHSS), time in days since stroke, and modified Rankin Scale (mRS).

Process of revising the MET-Home

In my previous study (see previous section), the original version of the MET-Home was completed by 25 stroke survivors and healthy controls. However, trends against completing items 1 ('Call a plumber and ask about the cost of a 1hr service call') and 12 ('Locate the cost of ordering a large two-item pizza for delivery') of the MET-Home were identified through both informal and formal feedback from participants. As a result, a decision was made to pragmatically revise the MET-Home during the study. The revision is reported in this section.

Initial item replacements I proposed were accepted as alternatives by the MET-Home authors. Item one was replaced by 'Call a membership service provider (e.g., gym, club, community centre etc.) and ask about the cost of a membership plan (can be daily, weekly + etc.).' Item 12 was revised due to non-recorded informal feedback from participants during MET administration where the examiner was unable to probe for more information. Participant comments included that they couldn't find the cost of a large two-item pizza for delivery as their village was too small/far out for any pizza place to deliver it. After testing, I discussed with the participants that they could still get a price online or through others, even if that place wouldn't deliver, but participants refused to accept this. In light of these anecdotal comments, item 12 was initially revised to 'What is the exact cost of ordering food for delivery to your home? (It can be takeaway food, or any

groceries etc)', but this led to questions from participants about whether delivery costs or food costs counted for accurate completion of the item. The final item 12 was revised to 'What is the exact cost of ordering some food to your home (inclusive of food costs; e.g., can be a takeaway, or food shopping etc)?'. Item 12 was felt by myself and the original MET-Home author to adequately cover the need to find a financial cost and be related to food.

Procedure

Data was collected from the homes of participants. Once participants had given their consent and demographic data had been gathered, the MET-Home was administered to all participants by the first author, with assistance from two research assistants when needed.

Data analysis

Data from my UK sample was added to the original Northern American MET-Home data. Initially, the UK and US sample characteristics were compared to pinpoint potential performance-impacting differences. Differences in performance between the original MET-Home and MET-Home revised were then outlined. The accuracy, omissions, and partial completions on the two MET versions were statistically contrasted using ANCOVA analysis, adjusting for age, education, mobility, and mRS. Feedback from Webb and Demeyere [270] (also previous section) specific to participants' experiences with either MET-Home version was briefly analyzed. For psychometrics, the internal consistency of the MET-Home in the UK sample was first inspected, given they were the sole group with MET-Home item level data. Healthy controls from both countries were subsequently merged into a single control group, and a similar approach was taken for stroke survivors. The new combined normative data of all healthy controls is showcased. The known-group discrimination of healthy controls from stroke survivors based on MET-Home performance was then assessed. To garner some convergent validity evidence of

the MET-Home, associations between the MET-Home and the MoCA, the Barthel Index, and the NEADL measures were also scrutinized.

Regarding power analyses, since the data was secondary, only sensitivity analyses could be undertaken to detect the noted effect sizes. Setting an alpha at .05 and with n= 98 individuals at a minimum, based on the f^2 statistic, the power was determined to range between 67-77%.

7.1.4 Results

Participants

In the UK, 51 stroke survivors and 58 healthy controls were recruited. The UK and US samples were statistically contrasted using t-tests and chi-squared tests, adhering to necessary assumptions. Multiple demographic variances in the samples for both healthy controls and stroke survivors were observed, predominantly related to age and ethnicity. The discrepancies in ethnicity can largely be attributed to the samples hailing from different nations. The North American sample, for both healthy controls and stroke survivors, was notably younger and had milder strokes.

A comparison and individual sample demographic summaries can be viewed in Table 7.7 in the appendix. The consolidated sample for analysis, encompassing the UK and North American samples, consisted of 66 stroke survivors and 78 healthy controls. The demographic data for this consolidated sample are outlined in Table 7.1. Any missing data is denoted in brackets, indicating the percentage relative to N per demographic. The majority of the missing data arises from varying demographic protocols between samples. No data imputation was executed for the entire sample.

Table 7.1: Summary of sample characteristics for all participants (23 stroke and healthy US participants, and 51 stroke and 58 healthy UK participants combined)

	Healthy aging		Stroke	
Characteristic	N (%)	Value	N (%)	Value
Age (M(SD))	77 (1%)	66.61 (13.17)	66 (0%)	64.79 (14.18)
Education (M(SD))	55 (29%)	15.58 (3.24)	43 (35%)	14.26 (3.72)
Handedness	55 (29%)	R: 89.09%; L: 7.27%; A: 3.64%;	43 (35%)	R: 86.05%; L: 11.63%; A: 2.33%;
Sex	77 (1%)	F: 50.65%; M: 49.35%;	66 (0%)	M: 53.03%; F: 46.97%;
Ethnicity	77 (1%)	White: English, Welsh, Scottish, Northern Irish or British: 67.53%; White: Any other white background: 20.78%; Other ethnic group: Any other ethnic group: 10.39%; Asian or Asian British: Indian: 1.3%;	66 (0%)	White: English, Welsh, Scottish, Northern Irish or British: 56.06%; White: Any other white background: 30.3%; Other ethnic group: Any other ethnic group: 12.12%; Black, Black British, Caribbean, or African: Any other Black, Black British, Caribbean, or African background: 1.52%;
NIHSS (Median(IQR))			62 (6%)	5 (2-7.75)
Days Since Stroke (M(SD))			65 (2%)	509.85 (793.72)
Stroke type			43 (35%)	ischaemic: 74.42%; intracerebral haemorrhage: 16.28%; TIA: 6.98%; subarachnoid haemorrhage: 2.33%;
Stroke side			63 (5%)	R: 49.21%; L: 41.27%; B: 9.52%;
Barthel Index (M(SD))	78 (0%)	19.86 (0.5)	66 (0%)	18.85 (2.19)
MoCA Total (M(SD))	78 (0%)	26.68 (2.3)	65 (2%)	23.83 (4.19)
Mobility	55 (29%)	Walks completely independently with no difficulty: 96.36%; Walks independently but with difficulty (may use aid, e.g., frame/stick): 3.64%; No sitting balance: 0%; Wheelchair + sitting balance: 0%; Walks with assistance of 2: 0%; Walks with assistance of 1: 0%; Undertermined: 0%;	43 (35%)	Walks completely independently with no difficulty: 46.51%; Walks independently but with difficulty (may use aid, e.g., frame/stick): 44.19%; Walks with assistance of 1: 9.3%; No sitting balance: 0%; Wheelchair + sitting balance: 0%; Walks with assistance of 2: 0%; Undertermined: 0%;
Modified Rankin Scale	55(29%)	0: 96.36%; 1: 1.82%; 2: 1.82%;	43(35%)	3: 34.88%; 2: 27.91%; 1: 23.26%; 0: 9.3%; 4: 4.65%;

Note. TIA = transient ischaemic attack.

Comparison between MET-Home and MET-Home revised

Only data from the UK sample was used initially, as this was the only dataset with item-level scoring. Upon visual inspection of item responses for both the original and revised MET-Home, it was observed that response patterns were similar (refer to Figure 7.1). For brevity, only the summary statistical results are presented for each set of comparisons regarding accuracy, omissions, and partial completions. In Figure 7.1, controls and stroke survivors on both versions of the MET-Home are detailed, and item-level accuracy, omissions, and partial completion data from the UK sample are provided.

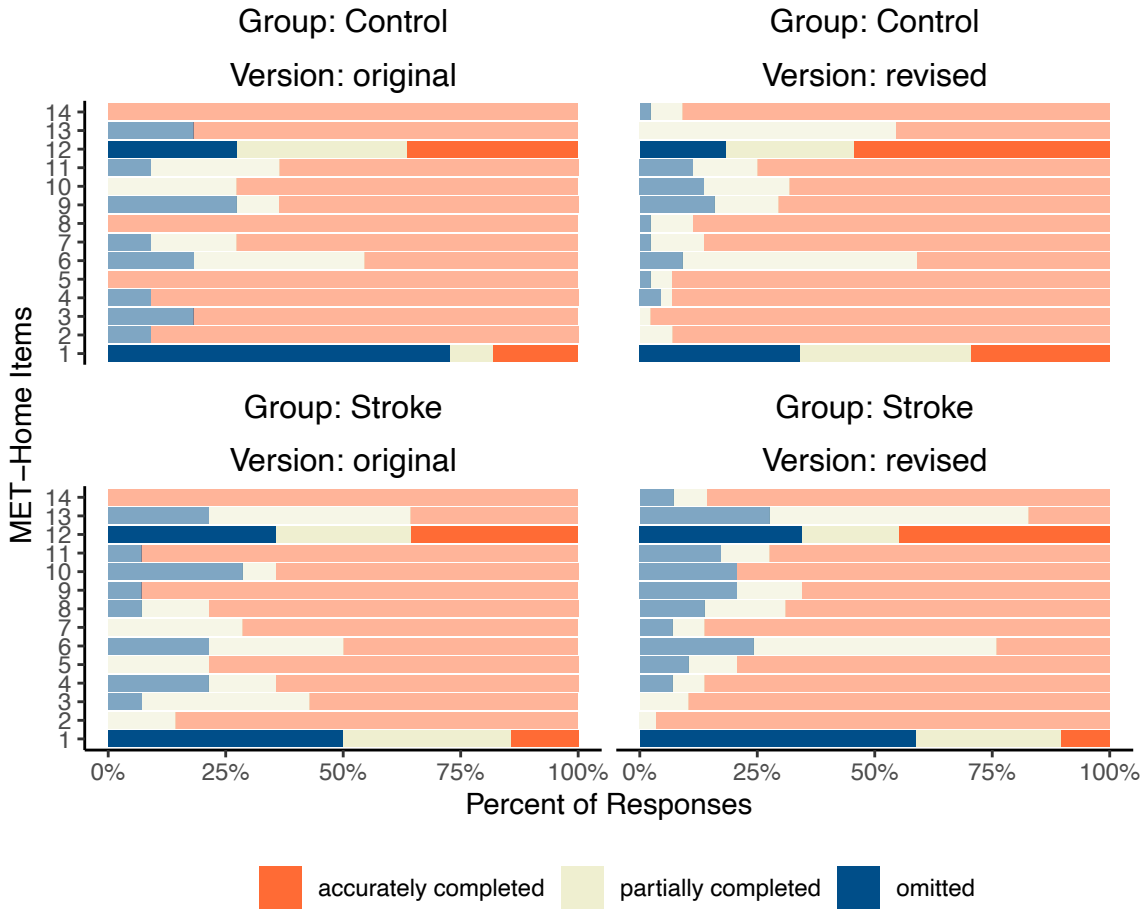


Figure 7.1: Item level responses for each item of the MET-Home, separated into original and revised versions (panel columns), and healthy controls and stroke survivors (panel rows). The changed items of 1 and 12 are highlighted in all panels.

In the original version for item 12 for all participants, there was an even spread

of accuracy (36%), omissions (32%), and partial completions (32%), whereas in the revised version 50.68% of participants were accurate, followed by 24.66% who omitted the task, and 24.68% who partially completed the task. A chi squared test revealed that there was no significant difference in versions on item 12 (x2(2)=1.61, p=.45) metrics. For item 1, I conducted the same analysis and found no significant difference in accuracy, omissions, or partial completions between the original and revised versions (x2(2)=1.95, p=.38). As presented next, as a whole test there were no statistically significant differences in performance.

Using covariates of age, education, mobility, and disability level from the UK sample, ANCOVA found no significant effect of version on accuracy scores, mean difference between original and revised = -0.07, $F(1, 91)= 0.76$, p=.38, generalized eta-squared = .01; nor for omission scores, mean difference between original and revised = 0.26, $F(1, 91)= 0.5$, p=.48, generalized eta-squared = 0; nor for partial completions, mean difference between original and revised = -0.25 $F(1, 91)= 0.24$, p=.62, generalized eta-squared = 0.

Additionally covarying for age, Barthel index, and sex within the combined country samples, ANCOVA analysis revealed no significant effect of version on accuracy scores, mean difference between original and revised = 0.71, $F(1, 138)= 0.02$, p=.89, generalized eta-squared = .01; nor omissions scores, mean difference between original and revised = -0.67, $F(1, 138)= 0.32$, p=.57 generalized eta-squared = 0; nor partial completions, mean difference between original and revised = -0.10, $F(1, 138)= 0.52$, p=.47, generalized eta-squared = 0.

For all stroke survivors covarying for stroke severity, days since stroke, sex, age, and Barthel index, the ANCOVA found no significant effect of version on accuracy scores, mean difference between original and revised = 0.44, $F(1, 54)= 1.62$, p=.21, generalized eta-squared = 0.03; nor for omissions, mean difference between original and revised = -1.17, $F(1, 54)= 0.64$, p=.43, generalized eta-squared = 0.01; but I did find a difference for partial completions mean difference between original and revised = 0.71, $F(1, 54)= 7.06$, p=.01, generalized eta-squared = 0.12.

In summary, there were largely no differences found between versions on the three main metrics from the MET-Home, except where stroke survivors had different amounts of partial completions. However, the significant result did not survive any correction for multiple comparisons.

Participant feedback by version

Differing participant numbers are reported in this manuscript from the MET-Home vs. OxMET study [270] due to the inclusion of more participants. 86 participants provided feedback. I examined those that made comments on item 1 of the MET-Home in either version, as this was changed in the revision, and no recorded comments pertained to the revised item 12. There were 14 participant comments on both the original and revised version that referred to item item 1. For the original MET-Home six of the eight participants simply did not want to complete item 1, four specified that calling was a waste of time (e.g., one control reported "*Phone the plumber is not good, it should be replaced by something else, as it would waste the plumber's time.*", and one specified they prefer to get the costs online).

For the revised version, one participant commented explicitly that they did not want to make a phone call, others commented on how external factors can influence their accuracy on item 1: one control participant commented "*making the call is tricky when the other person doesn't pick up*", another commented "*. . . and it is better when the things work [they tried to call 3 different membership services and only one picked up]*". A further four comments were made to the effect of the call being a waste of time (e.g., "*. . . dont want to waste time of membership services by calling.*").

In summary of feedback, not wanting to complete item 1 was not specific to version but to the modality of having to make a call at all.

Reliability

A split half reliability Cronbach's alpha with Spearman-Brown correction was conducted on all MET-Home accuracy item level data (14 items) for the UK sample as it was the only item data available. Internal consistency was different for the

original and revised versions of the MET-Home. The original version (n=23) had a reliability of a=.63, and the revised version (n=73) had a reliability of a=.80. A comparison of the Cronbach's alphas of each version via the '*cocron*' package in r [341] revealed that the reliabilities were not statistically different $x^2(1)$=3.24, p=.07. Across both versions combined reliability was high (a=.80) and was similar the originally reported reliability (e.g., a=.73) [126].

Known group discriminability

Next general group differences between healthy controls and stroke survivors across both samples were examined. Before this, it was established that there were no statistically differences in age ($t(133.98)$=.79, p=.43, d=.13) or education ($t(83.74)$=1.85, p=.07, d=.38) between the two groups. The analyses in Table 7.2 revealed that stroke survivors largely performed worse than healthy controls on many metrics; they had more partially completed tasks and more omitted tasks than healthy controls, and were less accurate. Additionally, stroke survivors were more inefficient and took longer to complete the tasks than healthy controls. There were no significant differences in planning time, number or frequency of rules broken, or number of passes by where a task could have been completed. I opted to use 5th centiles to classify impairment on each MET-Home metric, a common practice in neuropsychological screening [38, 50] and used with the OxMET previously [130]. Impairment cut offs are not typical in MET research, however, do benefit clinician interpretation of results.

Table 7.2: Wilcoxon comparisons between healthy controls and stroke suvivors on MET-Home metrics

MET-Home Metric	Healthy control median	Stroke survivor median	Wilcox statistic	p-value	95% CI lower	95% CI upper	effect size (r)
Accurately completed	11.5	10.0	3563.5	<.001	1	2	0.33
Partially completed	2.0	3.0	1813.5	<.001	-1	0	-0.26
Omitted	0.0	1.0	1782.5	<.001	-1	0	-0.28
Number of rules broken	2.0	3.0	1036.5	0.26	-1	0	-0.09
Frequency of rule breaks	6.0	8.0	2235.0	0.17	-3	0	-0.11
Number of passes	2.0	3.0	2117.0	0.06	-1	0	-0.16
Number of inefficiencies	0.0	0.0	2165.0	0.02	0	0	-0.19
Completion time	590.0	744.0	1874.5	0.01	-288	-50	-0.22
Planning time	113.5	131.5	2526.5	0.85	-44	25	-0.02

Normative reference data

In Table 7.3, I present normative data with 5th and 95th centile cut offs for impairment classification based on all healthy ageing MET-Home data. Accuracy is the only metric with a 5th centile, the rest use 95th centiles as they are error/time metrics. Initial analysis on demographics affecting MET-Home accuracy revealed that age was significantly associated with accuracy scores ($r(141)$= -.38, p<.001) and education was not significantly associated, with accuracy scores ($r(96)$= .14, p= .17). As such, I chose to compute separate cut offs for age groups. Age groups were determined statistically by equal quantiles using the 'split' base R function and 'cut2' function from the 'Hmisc' R package [252], this procedure maximises the quality of the cutoffs based upon the sample and reduces multiple comparisons to fish for differences in different age groups. New normative cut offs by age group can be developed in larger samples.

Comparison to general cognitive screen To assess how this specific executive dysfunction task stacked up against a broader global cognitive screen, age-adjusted impairment classifications of the main outcome metric of accuracy were compared with impairment on the MoCA across the entire sample. Data from the MoCA was available for 149 participants. Using a MoCA cutoff of less than 23 and considering this impairment classification as the benchmark for global cognitive impairment, the MET-Home accuracy showed 7% as "true" positives, 72% as "true" negatives, 8% as "false" positives, and 13% as "false" negatives. A sensitivity of 35% and a specificity of 90% were revealed.

This indicates that when compared to the MoCA, the MET-Home is highly specific in identifying those without global cognitive impairment. However, it lacks sensitivity in pinpointing those who do; many were deemed unimpaired by the MET-Home but showed cognitive impairment as per the MoCA. In the entire sample, the MoCA flagged 20% as impaired, while the MET-Home identified 15%. Due to the absence of subtask normative data for the MoCA and the lack of MoCA item-level data in both datasets, the generation of impairment classifications for the MoCA's

executive functioning section isn't possible. This means that it remains unclear if those deemed impaired by the MoCA, but not by the MET-Home, showcased impairments in domains other than executive functioning.

Table 7.3: Normative data from combined US and UK samples for each MET-Home metric, adjusted for quantile based age brackets.

	all		less than 62		62 to 74		greater than 74	
Metric	Median	Cut off	Median	Cut off	Median	Cut off	Median	Cut off
Accurately completed	12	7	12	9	12	6	10	6
Partially completed	2	5	1	3	2	6	2	6
Omitted	0	4	0	1	0	4	2	4
Number of rules broken	2	3	2	3	2	3	2	3
Frequency of rule breaks	6	15	4	9	6	15	8	18
Number of passes	2	6	0	5	2	6	2	7
Number of inefficiencies	0	2	0	1	0	1	0	2
Completion time	590	1032	600	780	549	959	600	1128
Planning time	114	337	60	298	132	219	162	386

Note. Percentile cut offs are all 95th centile, thus impairment is a score greater than the presented value, except for accurately completed tasks., which uses the 5th centile and scores below this are considered impaired.

Association of MET-Home to cognitive, functional, and ADL measures

The Pearson correlations between MET-Home accuracy and total scores from the MoCA, the Barthel index, and Nottingham Extended ADL measures are shown in Figure 7.2, while raw correlations are provided in Table 7.4 with all metrics. I found that MET-Home accuracy had significant correlations with the Barthel, MoCA, and NEADL scores. The strongest correlation was with MoCA, and all correlations surpassed the convergent validity benchmark of .30 [129]. Concerning time and error metrics, no significant associations were identified at a significance level below .001, except for omissions, which was correlated with all three measures, and partial omissions, which was correlated with the MoCA.

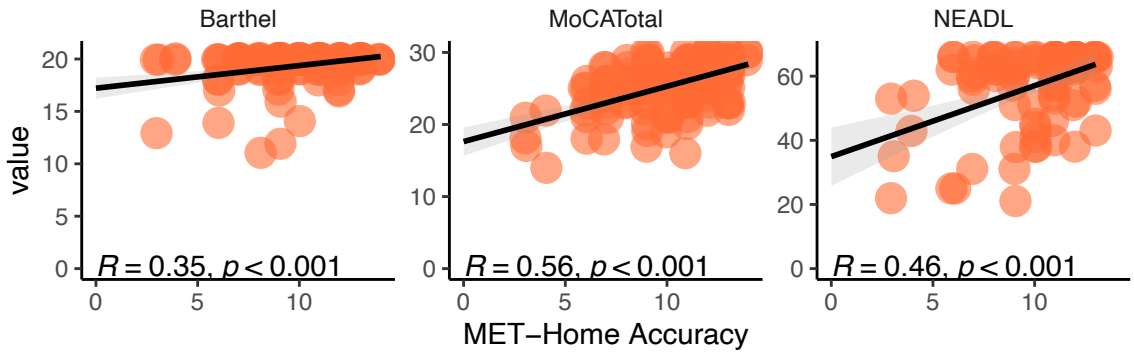


Figure 7.2: Associations betwen MET-Home accuracy and measures of functioning, cognition, and activities of daily living.

Table 7.4: Pearson correlations between MET-Home performance metrics and convergent measures inclusive of healthy controls and stroke survivors.

MET-Home metrics	Barthel Index	MoCA	NEADL
Planning time	-0.01	-0.15	-0.05
Time	-0.19*	0.03	-0.11
Accuracy	0.35**	0.56**	0.46**
Partial Omissions	-0.18*	-0.32**	-0.19
Omissions	-0.31**	-0.49**	-0.42**
Num. Rule Breaks	-0.08	-0.22*	-0.17
Freq. Rule Breaks	0.00	-0.12	0
Freq. Passes	0.05	-0.11	0.04
Inefficiencies	0.14	0.13	0.28*
Rooms Visited	-0.01	0.05	0.03

Note. '**' refers to significance below .001, '*' refers to significane below .05

7.1.5 Discussion

The aim of the current study was to present a revision of the MET-Home [126] and to provide reliability along with some convergent validity evidence for its use. Normative data, as a collation of UK and US sample MET-Home data, was also provided. It was discovered that there were limited statistically significant differences between the original and revised versions of the MET-Home on key metrics of accuracy, omissions, and partial completions; thus, version data was combined. Reliability, in terms of internal consistency, was found to be very high in the UK sample, which matched the high reliability identified in the original US sample [126].

Significant differences between age and education matched neurologically healthy control participants and survivors of stroke on accuracy, omissions, and partial completions, as well as time taken to complete the MET-Home and inefficiencies, were identified. However, group differences in planning time or number or frequency of rule breaks were not found.

Normative cut-offs were produced from the combined UK and US samples on all MET-Home metrics, and then impairment classification was compared with impairment on the Montreal Cognitive Assessment (MoCA) [37], an assessment of global cognitive impairment. It was found that the MET-Home exhibited very high specificity (90%) but very low sensitivity (30%) when compared to this global screen. The low sensitivity was investigated, revealing that 20% of the sample was impaired on the MoCA and 15% were impaired on the MET-Home. An inability to investigate what subtest (e.g., the trail making task) participants were impaired on one test and not the other to explain the difference is discussed later in the discussion. Finally, further evidence for the convergent validity of the MET-Home was established through correlation of MET-Home accuracy and global cognitive functioning (MoCA), functional independence [223], and instrumental activities of daily living [225]. It was determined that the MET-Home was highly related to these measures in accuracy and omissions and only partially for partial completions.

In comparison to the original investigation of the MET-Home [126], healthy controls and survivors of stroke were different on all MET-Home metrics except planning time (as found in the current study) and frequency of rule breaks. The original study did not compare impairment classification or MET-Home scores against the MoCA, but did correlate the MoCA with MET-Home metrics. The authors found that the MoCA was not statistically related to any MET-Home metrics, in contrast to the current findings of a moderate relationship between MET-Home accuracy and MoCA total score. In other MET research, it is usual to find group based differences, at least due to the presence of a brain injury in the patient group [124, 130, 268]. Nevertheless, it is important to establish that patient groups who are expected to perform worse, do so, to support the discriminability

and interpretation of the test results. In contrast to the results of the original MET-Home study, but in support of the current results, other MET literature has found associations between MET metrics and the MoCA [132, 268, 270, 279].

The MoCA, like the Barthel Index [223], is often used as an inclusion/exclusion criterion for participation in a study [270], whereby neurologically healthy controls are screened for cognitive impairment before participation (e.g., Antoniak et al. [134]. As the MoCA reflects general cognitive ability, previous research has used it as a covariate to establish the MET as a more specific test of executive dysfunction [237]. Previously, when controlling for MoCA score, the main effect of participant group on MET metrics became non-significant [237]. It is reasonable that there would be a difference in global cognitive ability between groups due to the presence of a brain injury, so it is unclear what can be interpreted by group differences becoming non-significant when accounting for cognitive abilities. It may only reveal a specific difference in a domain-specific ability (presumably executive function). These analyses did not include a specific executive assessment to say the difference in variance was due to executive dysfunction, so this cannot be concluded either. It must be noted then that the current investigation also did not include a specific executive function assessment, so distinctions between general and specific still cannot be made.

As the MET-Home is a specific assessment and the MoCA is more general, any comparisons which suggest no association may not reflect poorly on the psychometrics of the MET-Home, but to allude to the specific nature of the MET-Home and how it is not intended as a general screen. Without explicit research examining this association, no further conclusions can be made. Additionally, the MET-Home has both verisimilitude and veridicality in its assessment of executive dysfunction, which the MoCA does not have by design. The MET-Home, captures real life demands and activities that the MoCA does not, and by default, will explain different variance in a given participants performance.

Strengths and Limitations

A main limitation in the current study is the reliance on secondary data which has precluded some additional analyses. For instance, the US sample data did not have MET-Home item level data, which meant this data set could not be included in overall reliability analyses. Further, neither data set had item level MoCA data, preventing further investigation into the reason for a lack of sensitivity of the MET-Home to global cognitive impairment. Finally, there were demographic differences which could not be examined between US and UK samples as these data were not available in either the US or UK data sets. All data is reported where available.

Secondly, although the sample size of 78 is much larger than the original MET-Home paper sample of 23 healthy controls, this is still a small group from which to derive normative data. I had limitations on resources to dedicate to new normative data collection, and as such this is the maximum number that could be achieved. This lower sample size when including all participants may have affected the multiple comparison statistics found which were non-significant. However, these differences were so small in terms of effect size (new zero in some cases comparing versions with covariates in ANOVA) that even with larger samples the results are unlikely to become significant.

Clinical implications and Conclusions

Where previously, clinicians were tasked with diagnosing a problem, now, problems need to be identified in the context of how this affects a person's life going forward. In conjunction with the original paper by Burns et al. [126], the current study supports the use of either version of the MET-Home, but encourages the use of the revised version to have better acceptability by participants. The MET-Home is a reliable and homogeneous test with normative data for clinicians to reference. Clinicians can use the MET-Home for free to understand the possible impacts of executive dysfunction on everyday life. The MET-Home requires little formal training, and training videos are available on the test website (https://met-home.com/).

7.1.6 Wider Discussion

7.1.7 New normative data and concurrent validity for the MET-Home revised in english speaking healthy adults and stroke survivors

In this paper, I was the first to generate normative data for an in-person MET. Attempts have been made in the past to use 5th centile cut offs to classify participants as impaired compared to a control reference group (see example [125]), but the tables of cut offs have not been published. The simple reason for lack of normative data is that it is hard to provide standardised cut offs for ever evolving and changing tasks, such as the MET applied in different settings. MET tasks depends on location and tasks used. Thankfully, the MET-Home is standardised to mostly all typical homes, at least in the north western hemisphere. This means, normative data when can be applied for this region at least. Though cultural adaptions and translations are necessary when using any established test in another culture. This is the same as the OxMET, which is standardised and comes with automatically calculated impairment classifications based on evolving normative data. The difference with the OxMET is that it is designed to be culturally specific, and so it cannot apply to most people within the north western hemisphere. Three translations have been made for the OxMET and more are expected in the future. Furthermore, regarding normative data, the lack of normative data has affected how clinicians interpret performance. One qualitative examination of clinicians views on the utility of the MET, found that clinicians find it hard to know whether an error is normal or not and therefore how to interpret errors in the context of informing rehabilitation [342]. Indeed, one recent guideline for the MET called for a multi-site, multi-version, MET trial which could establish normative data for different METs to aid in clinician interpretation, after finding the same difficulties in interpreting scores [343]. One study from 2013 by Morrison et al. [257] stated that they were collecting new normative data for their MET-Revised version, but no such data have been released. I have emailed the author regarding this but never heard back. Due to the lack of normative data, the current study provides some much needed

reference. The normative sample is not the biggest and it is not the most diverse, but it is much better than the lack of information out there currently. With time, more normative data for MET-Home will be generated.

When comparing MET-Home to the MoCA, 79% of the time the MET-Home classified participants as impaired or spared in an identical way to the MoCA, which means 21% of participants were not classified in identical ways. This result does not mean the MET-Home is poor at classifying individuals. After all, MET-Home is a screening tool for the effects of executive dysfunction in everyday behaviours, it is not a global screening tool. The benefit of such analyses, however, is to establish the link between a short 10-minute global screening tool and the more specific MET-Home. The established link is a form of validity evidence in that performance on MET-Home is related to cognitive dysfunction, but explains some variance unexplained by this global screening tool. A shortcoming of the current study was the lack of executive functioning battery to measure executive functioning to establish if the unexplained variance was specific to dysfunction or not. Without this analyses I could not confirm this type of construct validity, though that was not the aim of the study. Regarding the results of associating MET-Home with other measures, the MET-Home was largely related to the MoCA score but also to ADLs in the form of basic and instrumental functioning. As a measure of the effects of executive dysfunction in everyday activities it is essential to establish that the MET-Home measures some kind of cognitive problem (established with the MoCA) and that it captures real life in some way (now establish through associations with functional outcomes). The missing piece is the specificity to executive function. This is the case for the OxMET too in many ways, as the original psychometric investigation of the OxMET did not do a thorough job of dissociating OxMET from other measures [130].

In summary, the MET-Home now has normative reference data stratified by age for clinicians to use when administering and interpreting MET-Home performance. This is much needed and will certainly fill a gap in clinician interpretation and translating performance to informing rehabilitation. Like other tests in this thesis,

the MET-Home also has strengths and weaknesses, and one such is the lack of evidence for specificity to executive dysfunction. To note for the MET-Home, no medical device regulation is necessary and it can be implemented in clinical practice as and when desired.

7.2 Tables and Futher data

Table 7.5: Scoring of MET-Home and OxMET.

Name	Description	References
Difficulties encountered when completing the task(s)	This theme surrounds issues encountered by participants when they were completing either MET task or when they were asked to complete it.	4
Issues with phone call element of MET-Home	This subtheme reflects the multiple comments specifically regarding making a phone call in the MET-Home.	21
Non-cognitive impairments make MET-Home harder	This subtheme reflects issues that were not cognition specific (where the MET's are supposed to measure cognition).	9
Stress and annoyance completing MET-Home	This subtheme reflects comments regarding the MET-Home inducing stress or strong negative emotions when completing the task.	5

Tablet vs phone	This subtheme reflects comments regarding the use of the OxMET on the phone. This was done for n=3 participants due to examiner error, whereby the examiner forgot to bring the computer tablet and had to use a phone instead.	3
Task was easy, simple, straightforward	This theme broadly covers the multiple neutral comments describing the METs as easy and fine.	40
Task was fun and interesting	This theme reflects comments regarding the interest and fun participants had when completing the METs.	38
Intellectual challenge	This subtheme reflects the challenge induced by the METs and how this was enjoyable.	19
Reflection of abilities after stroke	This theme reflects comments regarding participant abilities and how these have change since having a stroke.	3

Task requires planning or thinking before action	This theme reflects the multiple comments regarding how the METs require thinking before completion in order to complete the task well.	10
MET-Home	This is a specific category regarding all comments coded to do with the MET-Home specifically.	121
OxMET	This is a specific category regarding all comments coded to do with the OxMET specifically.	99
Attitude	Emotional valance of comments	0
Negative	All coded negative comments regardless of context	35
Neutral	All coded neutral comments regardless of context	54
Positive	All coded positive comments regardless of context	72

Table 7.6: Spearman's Rho correlations between subacute stroke survivor (<3 months post-stroke) OxMET performance metrics and functional outcome measures

OxMET metric	mRS	Barthel Index	TOMs total
Time	0.13	-0.05	-0.14
Accuracy	-0.30**	0.23*	0.27**
Omissions	0.04	-0.12	-0.21*
Partial Omissions	0.06	-0.08	-0.13
Freq. Rule Breaks	0.20*	-0.21	-0.21*
Perseveration	0.10	-0.20	-0.13
Commissions	0.21*	-0.16	-0.14
Total Errors	0.19*	-0.24*	-0.21*

Table 7.7: Summary of sample characteristics for all participants (healthy controls and stroke from the current study, and healthy controls and stroke from the original MET-Home), as well as statistical comparisons between them

	Current healthy controls		Original healthy controls			Current stroke		Original stroke		
Characteristic	N	Value	N	Value	Stat. comparison	N	Value	N	Value	Stat. comparison
Age (M(SD))	55 (0%)	70.73 (12.07)	22 (12%)	56.32 (9.86)	t= 5.42, p<.001*, d= 1.25	43 (0%)	69.1 (14.09)	23 (8%)	56.74 (10.55)	t= 4.02, p<.001*, d = 0.95
Education (M(SD))	55 (0%)	15.58 (3.24)				43 (0%)	14.26 (3.72)			
Handedness	55 (0%)	R: 89.09%; L: 7.27%; A: 3.64%;				43 (0%)	R: 86.05%; L: 11.63%; A: 2.33%;			
Sex	55 (0%)	F: 54.55%; M: 45.45%;	22 (12%)	M: 59.09%; F: 40.91%;	x2= 0.69, p=.41, V=0.12	43 (0%)	M: 58.14%; F: 41.86%;	23 (8%)	F: 56.52%; M: 43.48%;	x2= 0.77, p=.38, V=0.14
Ethnicity	55 (0%)	White: English, Welsh, Scottish, Northern Irish or British: 94.55%; White: Any other white background: 3.64%; Asian or Asian British: Indian: 1.82%;	22 (12%)	White: Any other white background: 63.64%; Other ethnic group: Any other ethnic group: 36.36%;	x2= 43.09, p<.001*, V=0.71	43 (0%)	White: English, Welsh, Scottish, Northern Irish or British: 86.05%; White: Any other white background: 9.3%; Black, Black British, Caribbean, or African: Any other Black, Black British, Caribbean, or African background: 2.33%; Other ethnic group: Any other ethnic group: 2.33%;	23 (8%)	White: Any other white background: 69.57%; Other ethnic group: Any other ethnic group: 30.43%;	x2= 26.8, p<.001*, V=0.56

Table 7.7: Summary of sample characteristics for all participants (healthy controls and stroke from the current study, and healthy controls and stroke from the original MET-Home), as well as statistical comparisons between them *(continued)*

Characteristic	N	Value	N	Value	Stat. comparison	N	Value	N	Value	Stat. comparison
NIHSS (Median(IQR))						39 (9%)	5 (3.5-9)	23 (8%)	2 (0-5)	t= 4.3, p<.001*, d= 0.98
Days Since Stroke (M(SD))						42 (2%)	539.64 (574.61)	23 (8%)	455.46 (1101.17)	t= 0.34, p=.73, d = 0.11
Stroke type						43 (0%)	ischaemic: 74.42%; intracerebral haemorrhage: 16.28%; TIA: 6.98%; subarachnoid haemorrhage: 2.33%;			
Stroke side						40 (7%)	R: 45%; L: 40%; B: 15%;	23 (8%)	R: 56.52%; L: 43.48%;	x2= 2.95, p=.23, V = 0.20
Barthel Index (M(SD))	55 (0%)	19.8 (0.59)	23 (8%)	20 (0)	t= -2.51, p=.01*, d=-0.40	43 (0%)	18.74 (2.19)	23 (8%)	19.04 (2.23)	t= -0.52, p=.60, d= -0.14
MoCA Total (M(SD))	55 (0%)	26.25 (2.43)	23(8%)	27.7 (1.61)	t= -3.08.p<.001, d= -0.65	42 (2%)	23.36 (4.5)	23(8%)	24.7 (3.46)	t= -1.34, p = .19, d= 0.32

Table 7.7: Summary of sample characteristics for all participants (healthy controls and stroke from the current study, and healthy controls and stroke from the original MET-Home), as well as statistical comparisons between them *(continued)*

Characteristic	N	Value	N	Value	Stat. comparison	N	Value	N	Value	Stat. comparison
Mobility	55 (0%)	Walks completely independently with no difficulty: 96.36%; Walks independently but with difficulty (may use aid, e.g., frame/stick): 3.64%; No sitting balance: 0%; Wheelchair + sitting balance: 0%; Walks with assistance of 2: 0%; Walks with assistance of 1: 0%; Undertermined: 0%;				43 (0%)	Walks completely independently with no difficulty: 46.51%; Walks independently but with difficulty (may use aid, e.g., frame/stick): 44.19%; Walks with assistance of 1: 9.3%; No sitting balance: 0%; Wheelchair + sitting balance: 0%; Walks with assistance of 2: 0%; Undertermined: 0%;			
Modified Rankin Scale	55(0%)	0: 96.36%; 1: 1.82%; 2: 1.82%;				43(0%)	3: 34.88%; 2: 27.91%; 1: 23.26%; 0: 9.3%; 4: 4.65%;			

Table 7.8: Scoring of MET-Home and OxMET.

MET version	Metric category	Metric name	Metric description and range
MET-Home	Time	Planning time	Time in seconds taken after the examiner starts the task until when the participant says they are ready to start the task, following guidance from MET-Home manual. Range 0 - inf.
		Time	Total time taken from when the participant says they have begun the task, to when they say they are finished or the researcher clarifies they are finished. Range 0 - inf.
	Accuracy	Accuracy	Number of tasks were completed in full and accurately. Ranges 0 – 14.
	Errors	Partial Completions	Number of tasks attempted but not completed or completed unsuccessfully. For example, a participant attempted to call someone for item 1 but they didn't get through, so they gave up and did not attempt it to completion. Ranges 0 – 14.
		Omissions	Number of tasks were not attempted or completed. Ranges 0 – 14.
		Number of Rule Breaks	Number of rules broken out of the 6 rules. If a rule is broken more than once, it only counts the first time. Ranges 0 – 6.
		Freq. Rule Breaks	Number of times any rule is broken throughout the task. For example, if a participant revisited a room more than once, each time is added to the frequency of rule breaks.

			Range 0 - inf.
	Inefficiencies	Freq. Passes	A pass is where an individual moves past an area where one or more of the tasks could have been attempted. Per each of the 14 items, only one pass counts. This is reflected on mostly when the task is completed.
			Range 0 – 14.
		Inefficiencies	Number of times participants became distracted or interrupted during the task. For example, if a participant got a call and did not end it quickly and get on with the task, or if they stopped to feed a pet etc, these count as inefficiencies.
			Range 0 - inf.
		Rooms Visited (or revisited)	When a room is visited it is recorded for tracking of room revisits. In the current study, this was dependent on residence type. For standardisation, if only one room is was used (it was multifunctional), researchers asserted to keep to each designated area and count it as different rooms. For example, if a participant had a large kitchen which had an attached living area and attached utility area, the researcher asked the participant to think of them as separate for the purpose of the study. This enabled there to be similar restrictions on each participant where this was possible. Room is up to researcher discretion.
			Range 0 - inf.
OxMET	Time	Time	Duration of time taken on OxMET

Accuracy	Accuracy	Total number of correctly completed tasks of 10 shops, where the presence of one or more errors deducts a point per shop, regardless of frequency of errors. Range -10 to 10.
Errors	Partial Omissions	The number of times a participant went into the correct shop but purchased a related item to the target but not the target item, and / or went into a shop to answer a question but never answered the question and /or answered a question without ever entering the shop related to the question. Range 0 – inf.
	Omissions	Number of times a participant failed to buy the target item. This tallies up the number of times the shop was entered and the target item not purchased, if they item had not already been purchased. For example, if a participant entered a target shop 10 items, and on each time they did something other than buy the target item (e.g., buy something else, or just leave) they scored an omission point. Range 0 – inf.
	Commissions	Number of times a participant purchased an unrelated item or entered a non-target shop. This also tallies for number of times a shop was entered, thus if a participant entered a target shop four times and each time they bought something either in addition to or instead of the target item, they scored a commission point. Range 0 – inf.

Perseveration	Number of revisits to the same shop at anypoint in the task. Range 0 – inf.
Freq. Rule Breaks	The number of times a rule is broken, no matter if it is the same rule. Range 0 – inf.
Total Errors	The frequency of errors made overall, irrespective of rule breaks. This score essentially tallies up partial omissions, omissions, commissions, and perseverations, where partial omissions are not a rule break. Range 0 – inf.

References

[1] Lynn Tondat Carter et al. "The Relationship of Cognitive Skills Performance to Activities of Daily Living in Stroke Patients". In: *American Journal of Occupational Therapy* 42.7 (July 1, 1988). Publisher: American Occupational Therapy Association, pp. 449–455. DOI: 10.5014/ajot.42.7.449. URL: https://ajot.aota.org/article.aspx?articleid=1881373 (visited on 11/11/2020).

[2] Nele Demeyere et al. "Long-term psychological consequences of stroke (OX-CHRONIC): A longitudinal study of cognition in relation to mood and fatigue after stroke: Protocol". In: *European Stroke Journal* 6.4 (2021). Publisher: SAGE Publications Sage UK: London, England, pp. 428–437.

[3] Deena Ebaid et al. "Mood and Cognitive Trajectories Over the First Year after Mild Ischemic Stroke". In: *Journal of Stroke and Cerebrovascular Diseases* 31.4 (Apr. 1, 2022), p. 106323. DOI: 10.1016/j.jstrokecerebrovasdis.2022.106323. URL: https://www.sciencedirect.com/science/article/pii/S1052305722000209 (visited on 02/28/2022).

[4] Bruce C. V. Campbell et al. "Ischaemic stroke". In: *Nature Reviews Disease Primers* 5.1 (Oct. 10, 2019). Number: 1 Publisher: Nature Publishing Group, pp. 1–22. DOI: 10.1038/s41572-019-0118-8. URL: https://www.nature.com/articles/s41572-019-0118-8 (visited on 08/12/2022).

[5] Valery L Feigin et al. "Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019". In: *The Lancet Neurology* 20.10 (Oct. 1, 2021), pp. 795–820. DOI: 10.1016/S1474-4422(21)00252-0. URL: https://www.sciencedirect.com/science/article/pii/S1474442221002520 (visited on 10/13/2022).

[6] Safi U. Khan et al. "Clinical and Economic Burden of Stroke Among Young, Midlife, and Older Adults in the United States, 2002-2017". In: *Mayo Clinic Proceedings: Innovations, Quality & Outcomes* 5.2 (Apr. 1, 2021), pp. 431–441. DOI: 10.1016/j.mayocpiqo.2021.01.015. URL: https://www.sciencedirect.com/science/article/pii/S2542454821000345 (visited on 10/13/2022).

[7] Y. X. Amy and Moira K. Kapral. "More people are surviving after acute stroke". In: *BMJ* 365 (2019). DOI: https://doi.org/10.1136/bmj.l2150.

[8] Ahmed Esmael, Mohammed Elsherief, and Khaled Eltoukhy. "Prevalence of cognitive impairment in acute ischaemic stroke and use of Alberta Stroke Programme Early CT Score (ASPECTS) for early prediction of post-stroke cognitive impairment". In: *Neurologia i Neurochirurgia Polska* 55.2 (2021), pp. 179–185.

[9] Andrea Kusec et al. "Long-term psychological outcomes following stroke: The OX-CHRONIC study". In: *BMC Neurology* 23 (2023). Publisher: Cold Spring Harbor Laboratory Press, p. 426.

[10] Niamh A. Merriman et al. "Addressing cognitive impairment following stroke: systematic review and meta-analysis of non-randomised controlled studies of psychological interventions". In: *BMJ open* 9.2 (2019). Publisher: British Medical Journal Publishing Group, e024429.

[11] Elise T. Milosevich et al. "Domain-specific cognitive impairment 6 months after stroke: the value of early cognitive screening". In: *medRxiv* (2023). Publisher: Cold Spring Harbor Laboratory Press, pp. 2023–06.

[12] Eithne Sexton et al. "Systematic review and meta-analysis of the prevalence of cognitive impairment no dementia in the first year post-stroke". In: *European Stroke Journal* 4.2 (June 1, 2019). Publisher: SAGE Publications, pp. 160–171. DOI: 10.1177/2396987318825484. URL: https://doi.org/10.1177/2396987318825484 (visited on 07/04/2022).

[13] S. Verstraeten, R. Mark, and M. Sitskoorn. "Motor and cognitive impairment after stroke: a common bond or a simultaneous deficit". In: *Stroke Research & Therapy* 1.1 (2016), p. 1.

[14] Stroke Association. *Lived Experience of Stroke Report*. 2018. URL: http://www.stroke.org.uk/sites/default/files/conferences/nisc/documents/lived_experience_of_stroke_chapter_1.pdf (visited on 11/11/2020).

[15] Joseph A. Mole and Nele Demeyere. "The relationship between early post-stroke cognition and longer term activities and participation: A systematic review". In: *Neuropsychological Rehabilitation* 30.2 (Mar. 2020), pp. 346–370. DOI: 10.1080/09602011.2018.1464934.

[16] N K J Oksala et al. "Cognitive impairment predicts poststroke death in long-term follow-up". In: *Journal of Neurology, Neurosurgery & Psychiatry* 80.11 (Nov. 1, 2009), pp. 1230–1235. DOI: 10.1136/jnnp.2009.174573. URL: https://jnnp.bmj.com/lookup/doi/10.1136/jnnp.2009.174573 (visited on 10/14/2020).

[17] T. Pohjasvaara et al. "Post-stroke depression, executive dysfunction and functional outcome". In: *European Journal of Neurology* 9.3 (2002). Publisher: Wiley Online Library, pp. 269–275. DOI: https://doi.org/10.1046/j.1468-1331.2002.00396.x.

[18] Hans Samuelsson et al. "Cognitive function is an important determinant of employment amongst young ischaemic stroke survivors with good physical recovery". In: *European Journal of Neurology* (2021). Publisher: Wiley Online Library. DOI: https://doi.org/10.1111/ene.15014.

[19] Ewa Lucka et al. "The Use of the ICF Classification Sheet to Assess Cognitive-Behavioral Disorders and Verbal Communication in Patients after Ischemic and Hemorrhagic Stroke during Rehabilitation". In: *International Journal of Environmental Research and Public Health* 19.19 (2022). Publisher: MDPI, p. 12127.

[20] Peter M Rothwell, Ale Algra, and Pierre Amarenco. "Medical treatment in acute and long-term secondary prevention after transient ischaemic attack and ischaemic stroke". In: *The Lancet* 377.9778 (May 14, 2011), pp. 1681–1692. DOI: 10.1016/S0140-6736(11)60516-3. URL: http://www.sciencedirect.com/science/article/pii/S0140673611605163 (visited on 09/28/2020).

[21] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, eds. *Standards for educational and psychological testing*. Washington (DC): American Educational Research Association, 2014.

[22] Samuel Messick. "Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning". In: *American Psychologist* 50 (1995). Place: US Publisher: American Psychological Association, pp. 741–749. DOI: 10.1037/0003-066X.50.9.741.

[23] Anne Anastasi. "Evolving concepts of test validation". In: *Annual review of Psychology* 37.1 (1986). Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, pp. 1–16.

[24] Michael T. Kane. "Validating the Interpretations and Uses of Test Scores". In: *Journal of Educational Measurement* 50.1 (2013). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jedm.12000, pp. 1–73. DOI: 10.1111/jedm.12000. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jedm.12000 (visited on 11/30/2021).

[25] Josiane Pawlowski et al. "A systematic review of validity procedures used in neuropsychological batteries." In: *Psychology & Neuroscience* 6.3 (2013). Publisher: Pontifícia Universidade Católica do Rio de Janeiro, p. 311. DOI: https://doi.org/10.3922/j.psns.2013.3.09.

[26] Emmanuelle Anthoine et al. "Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures". In: *Health and Quality of Life Outcomes* 12.1 (Dec. 9, 2014), p. 2. DOI: 10.1186/s12955-014-0176-2. URL: https://doi.org/10.1186/s12955-014-0176-2 (visited on 10/05/2021).

[27] Jeremy C. Hobart et al. "What sample sizes for reliability and validity studies in neurology?" In: *Journal of Neurology* 259.12 (Dec. 1, 2012), pp. 2681–2694. DOI: 10.1007/s00415-012-6570-y. URL: https://doi.org/10.1007/s00415-012-6570-y (visited on 10/05/2021).

[28] Jacob Cohen. "Statistical power analysis". In: *Current directions in psychological science* 1.3 (1992). tex.ids: RefWorks:doc:5c76a1a9e4b063d6170581e9, pp. 98–101. DOI: https://doi.org/10.1111/1467-8721.ep10768783.

[29] S Jones, S Carley, and M Harrison. "An introduction to power and sample size estimation". In: *Emergency Medicine Journal : EMJ* 20.5 (Sept. 2003), pp. 453–458. DOI: 10.1136/emj.20.5.453. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1726174/ (visited on 09/16/2020).

[30] Felix D. Schönbrodt and Marco Perugini. "At what sample size do correlations stabilize?" In: *Journal of Research in Personality* 47.5 (2013). tex.ids: schonbrodt_at_2013-1, pp. 609–612. DOI: https://doi.org/10.1016/j.jrp.2013.05.009.

[31] Bruno Kopp et al. "Errors on the Trail Making Test Are Associated with Right Hemispheric Frontal Lobe Damage in Stroke Patients". In: *Behavioural Neurology* 2015 (May 13, 2015). Publisher: Hindawi, e309235. DOI: 10.1155/2015/309235. URL: https://www.hindawi.com/journals/bn/2015/309235/ (visited on 10/05/2021).

[32] Katherine Arbuthnott and Janis Frank. "Trail making test, part B as a measure of executive control: validation using a set-switching paradigm". In: *Journal of clinical and experimental neuropsychology* 22.4 (2000). tex.ids: arbuthnott_trail_2000-1, pp. 518–528. DOI: https://doi.org/10.1076/1380-3395(200008)22:4;1-0;FT518.

[33] I. Sánchez-Cubillo et al. "Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities". In: *Journal of the International Neuropsychological Society* 15.3 (May 2009), pp. 438–450. DOI: 10.1017/S1355617709090626. URL: https://www.cambridge.org/core/product/identifier/S1355617709090626/type/journal_article (visited on 10/05/2021).

[34] Tom N. Tombaugh. "Trail Making Test A and B: normative data stratified by age and education". In: *Archives of clinical neuropsychology* 19.2 (2004). tex.ids: tombaugh_trail_2004-1, tombaugh_trail_2004-2, pp. 203–214. DOI: https://doi.org/10.1016/S0887-6177(03)00039-8.

[35] Norma Kabuba et al. "Effect of Age and Level of Education on Neurocognitive Impairment in HIV Positive Zambian Adults". In: *Neuropsychology* 32.5 (July 2018). Publisher: NIH Public Access, p. 519. DOI: 10.1037/neu0000438. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6296371/ (visited on 09/16/2020).

[36] Patricia A Lowe and Cecil R Reynolds. "Age, Gender, and Education May Have Little Influence on Error Patterns in the Assessment of Set-shifting and Rule Induction Among Normal Elderly". In: *Archives of Clinical Neuropsychology* 14.3 (Apr. 1, 1999), pp. 303–315. DOI: 10.1016/S0887-6177(98)00029-8. URL: http://www.sciencedirect.com/science/article/pii/S0887617798000298 (visited on 09/16/2020).

[37] Ziad S. Nasreddine et al. "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment". In: *Journal of the American Geriatrics Society* 53.4 (2005). tex.ids: RefWorks:doc:5bf03404e4b0b28cc72922c4, nasreddine_montreal_2005-1 publisher: Wiley Online Library, pp. 695–699. DOI: https://doi.org/10.1111/j.1532-5415.2005.53221.x.

[38] Nele Demeyere et al. "The Oxford Cognitive Screen (OCS): Validation of a stroke-specific short cognitive screening tool." In: *Psychological assessment* 27.3 (2015). tex.ids: RefWorks:doc:5bf12fd7e4b0b7882dc443e3, demeyere_oxford_2015-1, demeyere_oxford_2015-2, demeyere_oxford_2015-3, demeyere_oxford_2015-4, demeyere_oxford_2015-5, demeyere_oxford_2020, demeyere_oxford_2020-1, demeyere_oxford_2020-2 place: US publisher: American Psychological Association, p. 883. DOI: http://dx.doi.org/10.1037/pas0000082.

[39] Nele Demeyere et al. "Domain-specific versus generalized cognitive screening in acute stroke". In: *Journal of neurology* 263.2 (2016). Publisher: Springer, pp. 306–315. DOI: 10.1007/s00415-015-7964-4.

[40] Krista L Lanctôt et al. "Canadian Stroke Best Practice Recommendations: Mood, Cognition and Fatigue following Stroke, 6th edition update 2019". In: *International Journal of Stroke* 15.6 (Aug. 1, 2020). Publisher: SAGE Publications, pp. 668–688. DOI: 10.1177/1747493019847334. URL: https://doi.org/10.1177/1747493019847334 (visited on 10/17/2022).

[41] National Institute for and Health and Care Excellence. *Stroke Rehabilitation: Long term rehabilitation after stroke*. (NICE guideline (CG162). 2013. URL: https://www.nice.org.uk/guidance/cg162/evidence/full-guideline-pdf-190076509.

[42] Terence J. Quinn et al. "European Stroke Organisation and European Academy of Neurology joint guidelines on post-stroke cognitive impairment". In: *European Stroke Journal* 6.3 (Sept. 1, 2021). Publisher: SAGE Publications, pp. I–XXXVIII. DOI: 10.1177/23969873211042192. URL: https://doi.org/10.1177/23969873211042192 (visited on 07/04/2022).

[43] Intercollegiate Stroke Working Party. *National Clinical Guideline for Stroke for the UK and Ireland*. 2023. URL: www.strokeguideline.org.

[44] David McMahon, Clayton Micallef, and Terence J. Quinn. "Review of clinical practice guidelines relating to cognitive assessment in stroke". In: *Disability and Rehabilitation* 0.0 (Dec. 20, 2021). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09638288.2021.1980122, pp. 1–9. DOI: 10.1080/09638288.2021.1980122. URL: https://doi.org/10.1080/09638288.2021.1980122 (visited on 10/17/2022).

[45] Terence J. Quinn, Emma Elliott, and Peter Langhorne. "Cognitive and mood assessment tools for use in stroke". In: *Stroke* 49.2 (2018). Publisher: Am Heart Assoc, pp. 483–490.

[46] Anthony G. Rudd et al. "National clinical guideline for stroke: 2016". In: *Clinical Medicine* (2017). Publisher: Royal College of Physicians.

[47] Rosalind Lees et al. "Test Accuracy of Cognitive Screening Tests for Diagnosis of Dementia and Multidomain Cognitive Impairment in Stroke". In: *Stroke* 45.10 (Oct. 2014). Publisher: American Heart Association, pp. 3008–3018. DOI: 10.1161/STROKEAHA.114.005842. URL: https://www.ahajournals.org/doi/full/10.1161/strokeaha.114.005842 (visited on 10/17/2022).

[48] Marshal F. Folstein, Susan E. Folstein, and Paul R. McHugh. ""Mini-mental state": a practical method for grading the cognitive state of patients for the clinician". In: *Journal of psychiatric research* 12.3 (1975). tex.ids: RefWorks:doc:5bf03304e4b0b7882dc42b7f, folstein_mini-mental_1975-1 publisher: Pergamon, pp. 189–198.

[49] Yi-Jun Shen et al. "The use of MMSE and MoCA in patients with acute ischemic stroke in clinical". In: *International Journal of Neuroscience* 126.5 (May 3, 2016), pp. 442–447. DOI: 10.3109/00207454.2015.1031749. URL: http://www.tandfonline.com/doi/full/10.3109/00207454.2015.1031749 (visited on 09/06/2021).

[50] Nele Demeyere et al. "Introducing the tablet-based Oxford Cognitive Screen-Plus (OCS-Plus) as an assessment tool for subtle cognitive impairments". In: *Scientific Reports* 11.8000 (2021). DOI: https://doi.org/10.1038/s41598-021-87287-8. URL: https://psyarxiv.com/b2vgc/ (visited on 10/12/2020).

[51] P.S. Mathuranath et al. "A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia". In: *Neurology* 55.11 (Dec. 12, 2000), p. 1613. DOI: 10.1212/01.wnl.0000434309.85312.19. URL: http://n.neurology.org/content/55/11/1613.1.abstract.

[52] Miranda J. Say and Ciarán O'Driscoll. "Inter-rater variability in scoring of Addenbrooke's Cognitive Examination-Third Edition (ACE-III) protocols". In: *Applied Neuropsychology: Adult* (June 15, 2022), pp. 1–5. DOI: 10.1080/23279095.2022.2083964. URL: https://www.tandfonline.com/doi/full/10.1080/23279095.2022.2083964 (visited on 01/26/2023).

[53] Giuliana V. Zarrella et al. "Addenbrooke's Cognitive Examination–Third Edition Predicts Neuropsychological Test Performance". In: *The Journal of Neuropsychiatry and Clinical Neurosciences* (2022). Publisher: Am Neuropsych Assoc, appi–neuropsych.

[54] Ralph M. Reitan. "Trail making test results for normal and brain-damaged children". In: *Perceptual and motor skills* 33.2 (1971). Publisher: SAGE Publications Sage CA: Los Angeles, CA, pp. 575–581. DOI: https://doi.org/10.2466/pms.1971.33.2.575.

[55] YanHong Dong et al. "The Montreal Cognitive Assessment (MoCA) is superior to the Mini-Mental State Examination (MMSE) for the detection of vascular cognitive impairment after acute stroke". In: *Journal of the Neurological Sciences* 299.1 (Dec. 15, 2010). Publisher: Elsevier, pp. 15–18. DOI: 10.1016/j.jns.2010.08.051. URL: https://www.jns-journal.com/article/S0022-510X(10)00418-1/fulltext (visited on 03/01/2022).

[56] Sarah T. Pendlebury et al. "MoCA, ACE-R, and MMSE Versus the National Institute of Neurological Disorders and Stroke–Canadian Stroke Network Vascular Cognitive Impairment Harmonization Standards Neuropsychological Battery After TIA and Stroke". In: *Stroke* 43.2 (Feb. 2012). Publisher: American Heart Association, pp. 464–469. DOI: 10.1161/STROKEAHA.111.633586. URL: https://www.ahajournals.org/doi/full/10.1161/STROKEAHA.111.633586 (visited on 03/01/2022).

[57] Sarah T. Pendlebury et al. "Impact of Different Operational Definitions on Mild Cognitive Impairment Rate and MMSE and MoCA Performance in Transient Ischaemic Attack and Stroke". In: *Cerebrovascular Diseases* 36.5 (2013), pp. 355–362. DOI: 10.1159/000355496. URL: https://www.karger.com/Article/FullText/355496 (visited on 03/01/2022).

[58] Roy PC Kessels et al. "Regression-Based Normative Data for the Montreal Cognitive Assessment (MoCA) and Its Memory Index Score (MoCA-MIS) for Individuals Aged 18–91". In: *Journal of Clinical Medicine* 11.14 (2022). Publisher: MDPI, p. 4059.

[59] Louisa Burton and Sarah F. Tyson. "Screening for cognitive impairment after stroke: A systematic review of psychometric properties and clinical utility." In: *Journal of rehabilitation medicine* 47.3 (2015), pp. 193–203. URL: https://medicaljournalssweden.se/jrm/article/view/15528 (visited on 12/12/2023).

[60] Thomas J. Guilmette et al. "American Academy of Clinical Neuropsychology consensus conference statement on uniform labeling of performance test scores". In: *The Clinical Neuropsychologist* 34.3 (Apr. 2, 2020). Publisher: Routledge _eprint: https://doi.org/10.1080/13854046.2020.1722244, pp. 437–453. DOI: 10.1080/13854046.2020.1722244. URL: https://doi.org/10.1080/13854046.2020.1722244 (visited on 01/26/2023).

[61] Caroline M. Van Heugten, L. Walton, and U. Hentschel. "Can we forget the Mini-Mental State Examination? A systematic review of the validity of cognitive screening instruments within one month after stroke". In: *Clinical rehabilitation* 29.7 (2015). Publisher: SAGE Publications Sage UK: London, England, pp. 694–704.

[62] Godfred O. Boateng et al. "Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer". In: *Frontiers in Public Health* 6 (June 11, 2018). DOI: 10.3389/fpubh.2018.00149. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6004510/ (visited on 03/17/2021).

[63] T. B. Cumming et al. "Montreal Cognitive Assessment and Mini–Mental State Examination are both valid cognitive tools in stroke". In: *Acta Neurologica Scandinavica* 128.2 (2013). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ane.12084, pp. 122–129. DOI: 10.1111/ane.12084. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ane.12084 (visited on 09/06/2021).

[64] Katie Morris, Vicki Hacker, and Nadina Berrice Lincoln. "The validity of the Addenbrooke's Cognitive Examination-Revised (ACE-R) in acute stroke". In: *Disability and rehabilitation* 34.3 (2012). tex.ids: RefWorks:doc:5bf03a68e4b0efa59d30d0ca, pp. 189–195.

[65] Lucy C. Beishon et al. "Addenbrooke's Cognitive Examination III (ACE-III) and mini-ACE for the detection of dementia and mild cognitive impairment". In: *Cochrane Database of Systematic Reviews* 12 (2019). Publisher: John Wiley & Sons, Ltd.

[66] Ro Julia Robotham, Jens Oestergaard Riis, and Nele Demeyere. "A Danish version of the Oxford cognitive screen: a stroke-specific screening test as an alternative to the MoCA." In: *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition* 27.1 (Jan. 2020). Place: United States, pp. 52–65. DOI: 10.1080/13825585.2019.1577352.

[67] Adrian Wong et al. "Montreal cognitive assessment: one cutoff never fits all". In: *Stroke* 46.12 (2015). Publisher: Am Heart Assoc, pp. 3547–3550.

[68] Mauro Mancuso et al. "Using the Oxford Cognitive Screen to Detect Cognitive Impairment in Stroke Patients: A Comparison with the Mini-Mental State Examination." In: *Frontiers in neurology* 9 (2018), p. 101. DOI: 10.3389/fneur.2018.00101.

[69] Davide Crivelli et al. "The influence of language comprehension proficiency on assessment of global cognitive impairment following Acquired Brain Injury: A comparison between MMSE, MoCA and CASP batteries". In: *Applied Neuropsychology: Adult* (Aug. 21, 2021), pp. 1–6. DOI: 10.1080/23279095.2021.1966430. URL: https://www.tandfonline.com/doi/full/10.1080/23279095.2021.1966430 (visited on 09/06/2021).

[70] Julia Khaw et al. "Current update on the clinical utility of MMSE and MoCA for stroke patients in Asia: a systematic review". In: *International journal of environmental research and public health* 18.17 (2021). Publisher: MDPI, p. 8962. DOI: https://doi.org/10.3390/ijerph18178962.

[71] Pamela W. Duncan et al. "The stroke impact scale version 2.0: evaluation of reliability, validity, and sensitivity to change". In: *Stroke* 30.10 (1999). Publisher: Am Heart Assoc, pp. 2131–2140.

[72] Ching-Yi Wu et al. "Responsiveness, Minimal Clinically Important Difference, and Validity of the MoCA in Stroke Rehabilitation". In: *Occupational Therapy International* 2019 (Apr. 14, 2019). Publisher: Hindawi, e2517658. DOI: 10.1155/2019/2517658. URL: https://www.hindawi.com/journals/oti/2019/2517658/ (visited on 11/28/2023).

[73] Terence J. Quinn et al. "European Stroke Organisation and European Academy of Neurology joint guidelines on post-stroke cognitive impairment". In: *European Stroke Journal* (Oct. 8, 2021). Publisher: SAGE Publications, p. 23969873211042192. DOI: 10.1177/23969873211042192. URL: https://doi.org/10.1177/23969873211042192 (visited on 10/15/2021).

[74] Royal College of Physicians. *Clinical guide for the management of stroke patients during the coronavirus pandemic*. NHSE and NHSI, 2020.

[75] Porrselvi A. P et al. "The cultural and linguistic adaptation of the Oxford Cognitive Screen to Tamil". In: *Journal of the International Neuropsychological Society* 29.10 (Dec. 2023). Publisher: Cambridge University Press, pp. 964–971. DOI: 10.1017/S135561772300067X. URL: https://www.cambridge.org/core/journals/journal-of-the-international-neuropsychological-society/article/abs/cultural-and-linguistic-adaptation-of-the-oxford-cognitive-screen-to-tamil/5A1C6FD669FC8C328454F502F5034BFC (visited on 11/28/2023).

[76] Tania Corbí Pérez. "ADAPTACIÓN DE LA ESCALA OXFORD COGNITIVE SCREEN (OCS) EN POBLACIÓN ESPAÑOLA". In: (2016).

[77] M. D. García-Manzanares et al. "Cultural and linguistic adaptation of the Oxford Cognitive Screen to the Spanish population." In: *Neurologia (Barcelona, Spain)* (Jan. 21, 2020). Place: Spain. DOI: 10.1016/j.nrl.2019.11.004.

[78] Wen-Jun Hong et al. "Psychometric Properties of the Chinese (Putonghua) Version of the Oxford Cognitive Screen (OCS-P) in Subacute Poststroke Patients without Neglect." In: *BioMed research international* 2018 (2018), p. 6827854. DOI: 10.1155/2018/6827854.

[79] Hanne Huygelier et al. "The Dutch version of the Oxford Cognitive Screen (OCS-NL): normative data and their association with age and socio-economic status." In: *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition* 27.5 (Sept. 2020). Place: United States, pp. 765–786. DOI: 10.1080/13825585.2019.1680598.

[80] Anthony Pak-Hin Kong et al. "The Hong Kong version of the Oxford Cognitive Screen (HK-OCS): validation study for Cantonese-speaking chronic stroke survivors." In: *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition* 23.5 (Sept. 2016). Place: United States, pp. 530–548. DOI: 10.1080/13825585.2015.1127321.

[81] M. Mancuso et al. "Italian normative data for a stroke specific cognitive screening tool: the Oxford Cognitive Screen (OCS)." In: *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* 37.10 (Oct. 2016). Place: Italy, pp. 1713–1721. DOI: 10.1007/s10072-016-2650-6.

[82] Claudia Cristina Ferreira Ramos et al. "Oxford Cognitive Screen - Brazilian Portuguese version (OCS-Br) A pilot study." In: *Dementia & neuropsychologia* 12.4 (Dec. 2018), pp. 427–431. DOI: 10.1590/1980-57642018dn12-040014.

[83] Maria Shendyapina et al. "The Russian version of the Oxford Cognitive Screen: Validation study on stroke survivors." In: *Neuropsychology* 33.1 (Jan. 2019). Place: United States, pp. 77–92. DOI: 10.1037/neu0000491.

[84] Desirée Valera-Gran et al. "Validation of the Spanish version of the Oxford Cognitive Screen (S-OCS): psychometric properties of a short cognitive stroke-specific screening tool." In: *Clinical rehabilitation* 33.4 (Apr. 2019). Place: England, pp. 724–736. DOI: 10.1177/0269215518819046.

[85] Rudi Coetzer and Helen Bichard. "The challenges and opportunities of delivering clinical neuropsychology services during the Covid-19 crisis of 2020". In: *Cuadernos de Neuropsicología/Panamerican Journal of Neuropsychology* 14.1 (2020). DOI: 10.7714/CNPS/14.1.204.

[86] Sue Thomas et al. "Parkinson's disease and the COVID-19 pandemic: responding to patient need with nurse-led telemedicine". In: *British Journal of Neuroscience Nursing* 16.3 (2020). Publisher: MA Healthcare London, pp. 131–133. DOI: https://doi.org/10.12968/bjnn.2020.16.3.131.

[87] Sam S Webb, Eirini Kontou, and Nele Demeyere. "The COVID-19 pandemic altered the modality, but not the frequency, of formal cognitive assessment". In: *Disability and Rehabilitation* 0.0 (Aug. 16, 2021). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09638288.2021.1963855, pp. 1–9. DOI: 10.1080/09638288.2021.1963855. URL: https://doi.org/10.1080/09638288.2021.1963855 (visited on 09/06/2021).

[88] Sarah T. Pendlebury et al. "Telephone Assessment of Cognition After Transient Ischemic Attack and Stroke". In: *Stroke* 44.1 (Jan. 1, 2013). Publisher: American Heart Association, pp. 227–229. DOI: 10.1161/STROKEAHA.112.673384. URL: https://www.ahajournals.org/doi/10.1161/STROKEAHA.112.673384 (visited on 02/01/2022).

[89] Jared F. Benge and Andrew M. Kiselica. "Rapid communication: Preliminary validation of a telephone adapted Montreal Cognitive Assessment for the identification of mild cognitive impairment in Parkinson's disease". In: *The Clinical Neuropsychologist* 35.1 (Jan. 2, 2021), pp. 133–147. DOI: 10.1080/13854046.2020.1801848. URL: https://www.tandfonline.com/doi/full/10.1080/13854046.2020.1801848 (visited on 02/01/2022).

[90] Mindy J. Katz et al. "T-MoCA: A valid phone screen for cognitive impairment in diverse community samples". In: *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 13.1 (2021). _eprint: https://alz-journals.onlinelibrary.wiley.com/doi/pdf/10.1002/dad2.12144, e12144. DOI: https://doi.org/10.1002/dad2.12144. URL: https://alz-journals.onlinelibrary.wiley.com/doi/abs/10.1002/dad2.12144 (visited on 05/25/2021).

[91] Jason Brandt, Miriam Spencer, and Marshal Folstein. "The telephone interview for cognitive status". In: *Neuropsychiatry Neuropsychol Behav Neurol* 1.2 (1988), pp. 111–117.

[92] Lori A. Newkirk et al. "Validation of a 26-Point Telephone Version of the Mini-Mental State Examination". In: *Journal of Geriatric Psychiatry and Neurology* 17.2 (June 1, 2004). Publisher: SAGE Publications Inc STM, pp. 81–87. DOI: 10.1177/0891988704264534. URL: https://doi.org/10.1177/0891988704264534 (visited on 03/28/2022).

[93] Joseph T. Duffin et al. "Subtle memory and attentional deficits revealed in an Irish stroke patient sample using domain-specific cognitive tasks". In: *Journal of Clinical and Experimental Neuropsychology* 34.8 (Oct. 2012), pp. 864–875. DOI: 10.1080/13803395.2012.690368. URL: http://www.tandfonline.com/doi/abs/10.1080/13803395.2012.690368 (visited on 10/15/2021).

[94] Steven G. Faux et al. "Rehabilitation and education are underutilized for mild stroke and TIA sufferers". In: *Disability and Rehabilitation* 40.12 (June 5, 2018), pp. 1480–1484. DOI: 10.1080/09638288.2017.1295473. URL: https://www.tandfonline.com/doi/full/10.1080/09638288.2017.1295473 (visited on 10/17/2022).

[95] Marieke E. Verbraak et al. "Subtle Problems in Activities of Daily Living after a Transient Ischemic Attack or an Apparently Fully Recovered Non-disabling Stroke". In: *Journal of Stroke and Cerebrovascular Diseases* 21.2 (Feb. 1, 2012), pp. 124–130. DOI: 10.1016/j.jstrokecerebrovasdis.2010.05.012. URL: https://www.sciencedirect.com/science/article/pii/S1052305710001321 (visited on 10/17/2022).

[96] Sarah T. Pendlebury et al. "Underestimation of Cognitive Impairment by Mini-Mental State Examination Versus the Montreal Cognitive Assessment in Patients With Transient Ischemic Attack and Stroke: A Population-Based Study". In: *Stroke* 41.6 (June 2010), pp. 1290–1293. DOI: 10.1161/STROKEAHA.110.579888. URL: https://www.ahajournals.org/doi/10.1161/STROKEAHA.110.579888 (visited on 08/14/2020).

[97] S.N. McLennan et al. "Validity of the Montreal Cognitive Assessment (MoCA) as a Screening Test for Mild Cognitive Impairment (MCI) in a Cardiovascular Population". In: *Journal of Geriatric Psychiatry and Neurology* 24.1 (Mar. 1, 2011). Publisher: SAGE Publications Inc STM, pp. 33–38. DOI: 10.1177/0891988710390813. URL: https://doi.org/10.1177/0891988710390813 (visited on 10/28/2022).

[98] YanHong Dong et al. "Brief screening tests during acute admission in patients with mild stroke are predictive of vascular cognitive impairment 3–6 months after stroke". In: *Journal of Neurology, Neurosurgery & Psychiatry* 83.6 (June 2012), pp. 580–585. DOI: 10.1136/jnnp-2011-302070. URL: https://jnnp.bmj.com/lookup/doi/10.1136/jnnp-2011-302070 (visited on 10/28/2022).

[99] Nicole S. Schmid et al. "The extension of the German CERAD neuropsychological assessment battery with tests assessing subcortical, executive and frontal functions improves accuracy in dementia diagnosis". In: *Dementia and geriatric cognitive disorders extra* 4.2 (2014). Publisher: Karger Publishers, pp. 322–334.

[100] P. A. Osterrieth. "Le test de copie d'une figure complexe; contribution à l'étude de la perception et de la mémoire. [Test of copying a complex figure; contribution to the study of perception and memory.]" In: *Archives de Psychologie* 30 (1944). Place: Switzerland Publisher: Editions Médecine et Hygiène, pp. 206–356.

[101] B. Wilson, J. Cockburn, and P. Halligan. "Development of a behavioral test of visuospatial neglect." In: *Archives of physical medicine and rehabilitation* 68.2 (1987), pp. 98–102.

[102] Aaron T. Beck et al. "Beck depression inventory (BDI)". In: *Arch Gen Psychiatry* 4.6 (1961), pp. 561–571.

[103] Naomi Chaytor and Maureen Schmitter-Edgecombe. "The Ecological Validity of Neuropsychological Tests: A Review of the Literature on Everyday Cognitive Skills". In: *Neuropsychology Review* 13.4 (Dec. 2003), pp. 181–197. DOI: 10.1023/B:NERV.0000009483.91468.fb. URL: http://link.springer.com/10.1023/B:NERV.0000009483.91468.fb (visited on 11/02/2022).

[104] Robert K. Heaton and Mark G. Pendleton. "Use of Neuropsychological tests to predict adult patients' everyday functioning." In: *Journal of consulting and clinical psychology* 49.6 (1981). Publisher: American Psychological Association, p. 807.

[105] Paul W. Burgess et al. "The ecological validity of tests of executive function". In: *Journal of the international neuropsychological society* 4.6 (1998). tex.ids: RefWorks:doc:5bd60428e4b0954dddb19f63, pp. 547–558. DOI: https://doi.org/10.1017/S1355617798466037.

[106] Timothy Shallice and Paul Burgess. "Deficits in strategy application following frontal lobe damage in man". In: *Brain : a journal of neurology* 114.2 (1991). tex.ids: shallice_deficits_1991-1, shallice_deficits_1991-2 ISBN: 0006-8950 publisher: Oxford University Press, pp. 727–741. DOI: https://doi.org/10.1093/brain/114.2.727.

[107] David E. Tupper and Keith D. Cicerone. "Introduction to the neuropsychology of everyday life". In: *The neuropsychology of everyday life: Assessment and basic competencies*. Springer, 1990, pp. 3–18.

[108] Elizabeth R. Skidmore, Gail Eskes, and Amy Brodtmann. "Executive Function Poststroke: Concepts, Recovery, and Interventions". In: *Stroke* 54.1 (2023). Publisher: Am Heart Assoc, pp. 20–29.

[109] Sam Goldstein et al. "Introduction: A history of executive functioning as a theoretical and clinical construct". In: *Handbook of executive functioning*. tex.ids: goldstein_introduction_2014-1. Springer, 2014, pp. 3–12.

[110] Robert M. Bilder. "Executive control: balancing stability and flexibility via the duality of evolutionary neuroanatomical trends". In: *Dialogues in Clinical Neuroscience* 14.1 (Mar. 31, 2012). Publisher: Taylor & Francis _eprint: https://doi.org/10.31887/DCNS.2012.14.1/rbilder, pp. 39–47. DOI: 10.31887/DCNS.2012.14.1/rbilder. URL: https://doi.org/10.31887/DCNS.2012.14.1/rbilder (visited on 11/02/2022).

[111] A. R. Luria et al. "Impaired selectivity of mental processes in association with a lesion of the frontal lobe". In: *Neuropsychologia* 5.2 (May 1, 1967), pp. 105–117. DOI: 10.1016/0028-3932(67)90012-7. URL: https://www.sciencedirect.com/science/article/pii/0028393267900127 (visited on 11/02/2022).

[112] M. S. George and Sam Gilbert. "Mental Capacity Act (2005) assessments: Why everyone needs to know about the frontal lobe paradox". In: *The Neuropsychologist* 5.1 (2018), pp. 59–66.

[113] Simon Newstead et al. "The Paradox of the Frontal Lobe Paradox. A Scoping Review". In: *Frontiers in Psychiatry* 13 (July 22, 2022), p. 913230. DOI: 10.3389/fpsyt.2022.913230. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9353325/ (visited on 01/27/2023).

[114] Paul W. Burgess. "Strategy application disorder: the role of the frontal lobes in human multitasking". In: *Psychological Research* 63.3 (Aug. 18, 2000), pp. 279–288. DOI: 10.1007/s004269900006. URL: http://link.springer.com/10.1007/s004269900006 (visited on 12/11/2023).

[115] Paul W Burgess et al. "The cognitive and neuroanatomical correlates of multitasking". In: *Neuropsychologia* 38.6 (June 1, 2000), pp. 848–863. DOI: 10.1016/S0028-3932(99)00134-7. URL: https://www.sciencedirect.com/science/article/pii/S0028393299001347 (visited on 12/11/2023).

[116] Zhongmin Lin et al. "Trail Making Test Performance Using a Touch-Sensitive Tablet: Behavioral Kinematics and Electroencephalography". In: *Frontiers in Human Neuroscience* 15 (July 1, 2021), p. 663463. DOI: 10.3389/fnhum.2021.663463. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8281242/ (visited on 10/05/2021).

[117] Michael D. Franzen and Karen L. Wilhelm. "Conceptual foundations of ecological validity in neuropsychological assessment". In: *Ecological validity of neuropsychological testing* (1996), pp. 91–112.

[118] Anneliese A. Pontius and Bernard S. Yudowitz. "Frontal lobe system dysfunction in some criminal actions as shown in the narratives test." In: *Journal of Nervous and Mental Disease* (1980). Publisher: Lippincott Williams & Wilkins.

[119] Rebecca E. Ready, Laura Stierman, and Jane S. Paulsen. "Ecological validity of neuropsychological and personality measures of executive functions". In: *The Clinical Neuropsychologist* 15.3 (2001). Publisher: Taylor & Francis, pp. 314–323.

[120] January Durant et al. "Relationship between the Activities of Daily Living Questionnaire and the Montreal Cognitive Assessment". In: *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 4 (Jan. 1, 2016), pp. 43–46. DOI: 10.1016/j.dadm.2016.06.001. URL: https://www.sciencedirect.com/science/article/pii/S2352872916300306 (visited on 08/31/2023).

[121] José Carlos Millán-Calenti et al. "Cognitive impairment as predictor of functional dependence in an elderly sample". In: *Archives of Gerontology and Geriatrics* 54.1 (Jan. 1, 2012), pp. 197–201. DOI: 10.1016/j.archger.2011.02.010. URL: https://www.sciencedirect.com/science/article/pii/S0167494311000367 (visited on 08/31/2023).

[122] Miaoran Lin et al. "The Prognostic Value of Domain-Specific Cognitive Abilities Assessed by Chinese Version of Oxford Cognitive Screen on Determining ADLs Recovery in Patients with Post-Stroke Cognitive Impairment". In: *Evidence-Based Complementary and Alternative Medicine* 2022 (Sept. 6, 2022). Publisher: Hindawi, e1084901. DOI: 10.1155/2022/1084901. URL: https://www.hindawi.com/journals/ecam/2022/1084901/ (visited on 11/28/2023).

[123] Antonio Luigi Bisogno et al. "The Oxford cognitive screen (OCS) as an acute predictor of long-term functional outcome in a prospective sample of stroke patients". In: *Cortex* 166 (Sept. 1, 2023), pp. 33–42. DOI: 10.1016/j.cortex.2023.04.015. URL: https://www.sciencedirect.com/science/article/pii/S0010945223001181 (visited on 11/28/2023).

[124] Caroline Knight, Nick Alderman, and Paul W. Burgess. "Development of a simplified version of the multiple errands test for use in hospital settings". In: *Neuropsychological Rehabilitation* 12.3 (2002). tex.ids: knight_development_2002-1 ISBN: 0960-2011, pp. 231–255. DOI: 10.1080/09602010244000039.

[125] Nick Alderman et al. "Ecological validity of a simplified version of the multiple errands shopping test". In: *Journal of the International Neuropsychological Society* 9.1 (2003). ISBN: 1355-6177, pp. 31–44. DOI: 10.1017/S1355617703910046.

[126] Suzanne Perea Burns et al. "Development, Reliability, and Validity of the Multiple Errands Test Home Version (MET–Home) in Adults With Stroke". In: *American Journal of Occupational Therapy* 73.3 (2019). Publisher: American Occupational Therapy Association, 7303205030p1–7303205030p10. DOI: https://doi.org/10.5014/ajot.2019.027755.

[127] Diana Jovanovski et al. "Development of a novel, ecologically oriented virtual reality measure of executive function: The Multitasking in the City Test". In: *Applied Neuropsychology: Adult* 19.3 (2012). tex.ids: jovanovski_development_2012-1, pp. 171–182. DOI: https://doi.org/10.1080/09084282.2011.643955.

[128] Jamiahus Walton et al. "The Team Multiple Errands Test: A Platform to Evaluate Distributed Teams". In: *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*. CSCW '15: Computer Supported Cooperative Work and Social Computing. Vancouver BC Canada: ACM, Feb. 28, 2015, pp. 247–250. DOI: 10.1145/2685553.2699018. URL: https://dl.acm.org/doi/10.1145/2685553.2699018 (visited on 12/14/2023).

[129] Shlomit Rotenberg et al. "Measurement Properties of the Multiple Errands Test: A Systematic Review". In: *Archives of Physical Medicine and Rehabilitation* 101.9 (2020). tex.ids: rotenberg_measurement_2020-1, pp. 1628–1842. DOI: https://doi.org/10.1016/j.apmr.2020.01.019.

[130] Sam S Webb et al. "The Oxford Digital Multiple Errands Test (OxMET): Validation of a Simplified Computer Tablet Based Multiple Errands Test". In: *Neuropsychological Rehabilitation* (2021). DOI: 10.1080/09602011.2020.1862679.

[131] Suzanne Perea Burns et al. "In-home contextual reality: a qualitative analysis using the Multiple Errands Test Home Version (MET-Home)". In: *Neuropsychological Rehabilitation* 30.5 (2020). Publisher: Taylor & Francis, pp. 787–801.

[132] Frank Ho-yin Lai, Elaine Wai-hung Yan, and Kathy Ka-ying Yu. "Home-based evaluation of executive function (Home-MET) for older adults with mild cognitive impairment". In: *Archives of Gerontology and Geriatrics* 87 (Mar. 1, 2020), p. 104012. DOI: 10.1016/j.archger.2020.104012. URL: https://www.sciencedirect.com/science/article/pii/S0167494320300054 (visited on 03/16/2021).

[133] Deirdre R. Dawson et al. "Further Development of the Multiple Errands Test: Standardized Scoring, Reliability, and Ecological Validity for the Baycrest Version". In: *Archives of Physical Medicine and Rehabilitation* 90.11 (2009). tex.ids: dawson_further_2009-1 ISBN: 0003-9993, S41–S51. DOI: 10.1016/j.apmr.2009.07.012.

[134] Kristen Antoniak et al. "Developing and Validating a Big-Store Multiple Errands Test". In: *Frontiers in Psychology* 10 (2019). Publisher: Frontiers. DOI: 10.3389/fpsyg.2019.02575. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02575/full (visited on 09/29/2020).

[135] Pietro Cipresso et al. "Virtual multiple errands test (VMET): a virtual reality-based tool to detect early executive functions deficit in Parkinson's disease". In: *Frontiers in behavioral neuroscience* 8 (2014). Publisher: Frontiers, p. 405. DOI: https://doi.org/10.3389/fnbeh.2014.00405.

[136] D. Rand et al. "The virtual mall: A functional virtual environment for stroke rehabilitation". In: *Annual Review of Cybertherapy and Telemedicine: A decade of VR* 3 (2005), pp. 193–198.

[137] Simona Raspelli et al. "Validating the neuro VR-based virtual version of the multiple errands test: preliminary results". In: *Presence: Teleoperators and Virtual Environments* 21.1 (2012). tex.ids: RefWorks:doc:5c76acc0e4b0061ca5b7721a, pp. 31–42.

[138] Tracey Quinn. "Development of novel computerised tools to assess memory and planning problems in people with brain injury". PhD thesis. University of Glasgow, 2014.

[139] Barbara A. Wilson et al. "Behavioural assessment of the dysexecutive syndrome". In: *Methodology of frontal and executive function* 239 (1997). Publisher: Psychology Press Hove, p. 250.

[140] Debbie Rand and Noomi Katz. "Training multitasking in a virtual supermarket: a novel intervention after stroke." In: *American Journal of Occupational Therapy* 63.5 (2009). tex.ids: rand_training_2009-1. DOI: https://doi.org/10.5014/ajot.63.5.535simm.

[141] Neta Erez et al. "Comparing performance within a virtual supermarket of children with traumatic brain injury to typically developing children: A pilot study". In: *OTJR: occupation, participation and health* 33.4 (2013). Publisher: SAGE Publications Sage CA: Los Angeles, CA, pp. 218–227.

[142] Sayaka Okahashi et al. "A virtual shopping test for realistic assessment of cognitive function". In: *Journal of NeuroEngineering and Rehabilitation* 10.1 (June 18, 2013), p. 59. DOI: 10.1186/1743-0003-10-59. URL: https://doi.org/10.1186/1743-0003-10-59.

[143] Barbara Wilson et al. "The development and validation of a test battery for detecting and monitoring everyday memory problems". In: *Journal of Clinical and Experimental Neuropsychology* 11.6 (1988), pp. 855–870. URL: https://www.tandfonline.com/doi/abs/10.1080/01688638908400940?casa_token=eKWl2ZrursUAAAAA:6tdtn2jfFAmoy4OoV4OgbacCfnjjg9Le59TTxgu9DzyJ-oRQv3Jkg8AP09qJNDZebRTVgkbQQA (visited on 09/18/2020).

[144] Natalie Waugh. "Self-report of the the Young, Middle-aged, Young-old and Old-old Individuals on Prospective Memory Functioning". PhD thesis. Griffith University, 1999.

[145] M. Powell Lawton and Elaine M. Brody. "Assessment of older people: self-maintaining and instrumental activities of daily living". In: *The gerontologist* 9.3 (1969). Publisher: The Gerontological Society of America, pp. 179–186.

[146] Vanessa L Hanberg, Diane E MacKenzie, and Brenda K Merritt. "Scoping review of the Multiple Errands Test: Is it relevant to youths with acquired brain injury?" In: *British Journal of Occupational Therapy* 81.12 (Dec. 1, 2018). Publisher: SAGE Publications Ltd STM, pp. 673–686. DOI: 10.1177/0308022618791714. URL: https://doi.org/10.1177/0308022618791714 (visited on 09/18/2020).

[147] N. Titov and Robert G. Knight. "A computer-based procedure for assessing functional cognitive skills in patients with neurological injuries: the virtual street". In: *Brain Injury* 19.5 (2005). Publisher: Taylor & Francis, pp. 315–322.

[148] Panagiotis Kourtesis et al. "Validation of the Virtual Reality Everyday Assessment Lab (VR-EAL): An Immersive Virtual Reality Neuropsychological Battery with Enhanced Ecological Validity". In: *Journal of the International Neuropsychological Society* (Aug. 10, 2020), pp. 1–16. DOI: 10.1017/S1355617720000764. URL: https://www.cambridge.org/core/product/identifier/S1355617720000764/type/journal_article (visited on 11/04/2020).

[149] Anna S. Law, Robert H. Logie, and David G. Pearson. "The impact of secondary tasks on multitasking in a virtual environment". In: *Acta Psychologica* 122.1 (2006). tex.ids: law_impact_2006-1, pp. 27–44.

[150] Panagiotis Kourtesis et al. "Guidelines for the Development of Immersive Virtual Reality Software for Cognitive Neuroscience and Neuropsychology: The Development of Virtual Reality Everyday Assessment Lab (VR-EAL), a Neuropsychological Test Battery in Immersive Virtual Reality". In: *Frontiers in Computer Science* 1 (2020). Publisher: Frontiers. DOI: 10.3389/fcomp.2019.00012. URL: https://www.frontiersin.org/articles/10.3389/fcomp.2019.00012/full?fbclid=IwAR1h3pBNrAJJ04yq0OevYNmLtifFKac4rCbvaYNjE11MwlXZY5pcqQDLsmQ (visited on 11/04/2020).

[151] Tom Manly et al. "Rehabilitation of executive function: facilitation of effective goal management on complex tasks using periodic auditory alerts". In: *Neuropsychologia* 40.3 (Jan. 1, 2002), pp. 271–281. DOI: 10.1016/S0028-3932(01)00094-X. URL: http://www.sciencedirect.com/science/article/pii/S002839320100094X (visited on 11/04/2020).

[152] Andrea L. Zartman et al. "The Pillbox Test: an ecological measure of executive functioning and estimate of medication management abilities". In: *Archives of Clinical Neuropsychology* 28.4 (2013). Publisher: Oxford University Press, pp. 307–319. DOI: https://doi.org/10.1093/arclin/act014.

[153] R Core Team. *R: A language and environment for statistical computing*. manual. Vienna, Austria: R Foundation for Statistical Computing, 2022. URL: https://www.R-project.org/.

[154] Ulrik Lyngs. *oxforddown: An oxford university thesis template for R markdown*. 2019. DOI: 10.5281/zenodo.3484681. URL: https://github.com/ulyngs/oxforddown.

[155] Francois-Xavier Lescure et al. "Clinical and virological data of the first cases of COVID-19 in Europe: a case series". In: *The Lancet Infectious Diseases* (2020). Publisher: Elsevier.

[156] Greg Dropkin. "COVID-19 UK Lockdown Forecasts and R0". In: *Frontiers in Public Health* 8 (2020). Publisher: Frontiers, p. 256.

[157] Jodie E Chapman et al. "The use of videoconferencing in clinical neuropsychology practice: A mixed methods evaluation of neuropsychologists' experiences and views". In: *Australian Psychologist* (2020). _eprint: https://aps.onlinelibrary.wiley.com/doi/pdf/10.1111/ap.12471, pp. 1–16. DOI: 10.1111/ap.12471. URL: https://aps.onlinelibrary.wiley.com/doi/abs/10.1111/ap.12471 (visited on 10/30/2020).

[158] Natalie K. Bradford, Liam J. Caffery, and Anthony C. Smith. "Awareness, experiences and perceptions of telehealth in a rural Queensland community". In: *BMC Health Services Research* 15 (Sept. 28, 2015). DOI: 10.1186/s12913-015-1094-7. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4587917/ (visited on 12/15/2020).

[159] Mili Parikh et al. "Consumer acceptability of brief videoconference-based neuropsychological assessment in older individuals with and without cognitive impairment". In: *The Clinical Neuropsychologist* 27.5 (2013). Publisher: Taylor & Francis, pp. 808–817. DOI: https://doi.org/10.1080/13854046.2013.791723.

[160] Nathan C. Hantke and Christine Gould. "Examining Older Adult Cognitive Status in the Time of COVID-19". In: *Journal of the American Geriatrics Society* 68.7 (2020). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jgs.16514, pp. 1387–1389. DOI: https://doi.org/10.1111/jgs.16514. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jgs.16514 (visited on 12/08/2020).

[161] Professor Gary A Ford et al. "Adapting stroke services during the COVID-19 pandemic: an implementation guide". In: (2020), p. 57.

[162] NHS London Clinical Networks. *Guidance on remote working for memory services during COVID-19*. 2020. URL: https://www.rcpsych.ac.uk/docs/default-source/members/faculties/old-age/guidance-on-remote-working-for-memory-services-during-covid-19.pdf?sfvrsn=ef9b27a9_2.

[163] NHS England. *Statistics » Dementia Assessment and Referral 2020-21*. Dementia Assessment and Referral 2020-21. Sept. 2020. URL: https://www.england.nhs.uk/statistics/statistical-work-areas/dementia/_dementia-assessment-and-referral-2020-21/ (visited on 09/04/2020).

[164] Division of Neuropsychology. *Division of Neuropsychology Professional Standards Unit Guidelines to colleagues on the use of Tele-neuropsychology*. 2020. URL: https://www.bps.org.uk/sites/www.bps.org.uk/files/Member%20Networks/Divisions/DoN/DON%20guidelines%20on%20the%20use%20of%20tele-neuropsychology%20%28April%202020%29.pdf.

[165] Timothy W. Brearly et al. "Neuropsychological test administration by videoconference: A systematic review and meta-analysis". In: *Neuropsychology Review* 27.2 (2017). Publisher: Springer, pp. 174–186.

[166] Natalie A. Phillips et al. "Special Issues on Using the Montreal Cognitive Assessment for telemedicine Assessment During COVID-19". In: *J Am Geriatr Soc* (2020), pp. 19–21.

[167] MoCA Test Inc. *Remote MoCA Testing*. 2020. URL: https://mailchi.mp/mocatest/remote-moca-testing?e=bbeb81559c (visited on 09/04/2020).

[168] Leslie M. Guidotti Breting et al. "2020 COVID-19 American Academy of Clinical Neuropsychology (AACN) Student Affairs Committee survey of neuropsychology trainees". In: *The Clinical Neuropsychologist* 34.7 (Nov. 16, 2020), pp. 1284–1313. DOI: 10.1080/13854046.2020.1809712. URL: https://www.tandfonline.com/doi/full/10.1080/13854046.2020.1809712 (visited on 11/29/2023).

[169] David E Marra et al. "Initial changes in neuropsychologists clinical practice during the COVID-19 pandemic: A survey study". In: (), p. 17.

[170] Dustin B Hammers et al. "A survey of international clinical teleneuropsychology service provision prior to and in the context". In: Clinical Neuropsychology in the Time of COVID-19 34.17 (2020), pp. 1267–1283.

[171] Joseph P. Simmons, Leif D. Nelson, and Uri Simonsohn. "A 21 word solution". In: *Available at SSRN 2160588* (2012). DOI: https://dx.doi.org/10.2139/ssrn.2160588.

[172] G Eysenbach. "Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES)". In: *J Med Internet Res* 6.3 (), e34.

[173] Stephane Champely et al. "pwr: Basic functions for power analysis". In: *R package version* 1.1 (2015), p. 665.

[174] R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, 2021. URL: https://www.R-project.org/.

[175] Jisc. *JISC Online Surveys*. 2020. URL: https://www.onlinesurveys.ac.uk/ (visited on 09/11/2020).

[176] Eneida Mioshi et al. "The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening". In: *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences* 21.11 (2006). tex.ids: RefWorks:doc:5bf03974e4b0e21eda416ede, pp. 1078–1085.

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

[178] Christopher Randolph et al. "The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity". In: *Journal of clinical and experimental neuropsychology* 20.3 (1998). Publisher: Taylor & Francis, pp. 310–319.

[179] David Wechsler. "Wechsler Adult Intelligence Scale". In: *Archives of Clinical Neuropsychology* (1955).

[180] David Wechsler. "A standardized memory scale for clinical use". In: *The Journal of Psychology* 19.1 (1945). Publisher: Taylor & Francis, pp. 87–95.

[181] David Wechsler. *Wechsler Test of Adult Reading: WTAR.* Psychological Corporation, 2001.

[182] Justin Raymer. *Oxford Cognigive Screen remote administration version now available!* Oxford University Innovation. 2020. URL: https://innovation.ox.ac.uk/news/ocs-remote-administration/ (visited on 02/01/2022).

[183] Stephen R. Rapp et al. "Validation of a Cognitive Assessment Battery Administered over the Telephone". In: *Journal of the American Geriatrics Society* 60.9 (2012). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532-5415.2012.04111.x, pp. 1616–1623. DOI: 10.1111/j.1532-5415.2012.04111.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1532-5415.2012.04111.x (visited on 07/19/2022).

[184] A. C. Laska et al. "Aphasia in acute stroke and relation to outcome". In: *Journal of Internal Medicine* 249.5 (May 2001), pp. 413–422. DOI: 10.1046/j.1365-2796.2001.00812.x.

[185] Margaret J. Moore et al. "Recovery of Visuospatial Neglect Subtypes and Relationship to Functional Outcome Six Months After Stroke". In: *Neurorehabilitation and Neural Repair* 35.9 (Sept. 1, 2021). Publisher: SAGE Publications Inc STM, pp. 823–835. DOI: 10.1177/15459683211032977. URL: https://doi.org/10.1177/15459683211032977 (visited on 02/28/2022).

[186] Tim Coltman et al. "Formative versus reflective measurement models: Two applications of formative measurement". In: *Journal of Business Research.* Formative Indicators 61.12 (Dec. 1, 2008), pp. 1250–1262. DOI: 10.1016/j.jbusres.2008.01.013. URL: https://www.sciencedirect.com/science/article/pii/S0148296308000155 (visited on 07/01/2022).

[187] Glyn W. Humphreys et al. "Cognitive Function in Low-Income and Low-Literacy Settings: Validation of the Tablet-Based Oxford Cognitive Screen in the Health and Aging in Africa: A Longitudinal Study of an INDEPTH Community in South Africa (HAALSI)". In: *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 72.1 (2017). tex.ids: RefWorks:doc:5beefbb1e4b0b7882dc404ab, humphreys_cognitive_2017-1 ISBN: 1079-5014, pp. 38–50. DOI: 10.1093/geronb/gbw139.

[188] Erik von Elm et al. "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies". In: *The Lancet* 370.9596 (Oct. 20, 2007). Publisher: Elsevier, pp. 1453–1457. DOI: 10.1016/S0140-6736(07)61602-X. URL: https://doi.org/10.1016/S0140-6736(07)61602-X (visited on 12/20/2021).

[189] Clara Margarethe Hübner et al. "Validation of the Oxford Cognitive Screen-Plus: Tablet-based cognitive testing in patients with amnestic mild cognitive impairment". In: *Alzheimer's & Dementia* 16 (S6 2020). _eprint: https://alz-journals.onlinelibrary.wiley.com/doi/pdf/10.1002/alz.042557, e042557. DOI: 10.1002/alz.042557. URL: https://alz-journals.onlinelibrary.wiley.com/doi/abs/10.1002/alz.042557 (visited on 09/06/2021).

[190] The MathWorks Inc. *MATLAB and statistics toolbox*. Natick, MA, United States.

[191] Esther Van Den Berg et al. "The Brixton Spatial Anticipation Test as a test for executive function: validity in patient groups and norms for older adults". In: *Journal of the International Neuropsychological Society* 15.5 (2009). tex.ids: berg_brixton_2009-1, pp. 695–703.

[192] Allison A. M. Bielak et al. "Performance on the Hayling and Brixton tests in older adults: Norms and correlates". In: *Archives of Clinical Neuropsychology* 21.2 (2006). tex.ids: RefWorks:doc:5be158bbe4b0df4346c104aa ISBN: 0887-6177, pp. 141–149. DOI: 10.1016/j.acn.2005.08.006.

[193] Paul W Burgess and Tim Shallice. *The hayling and brixton tests*. tex.ids: burgess_hayling_1997-1. Thames Valley Test Company, 1997.

[194] Nancy Helm-Estabrooks. *Cognitive linguistic quick test: CLQT*. Psychological Corporation, 2001.

[195] Philip Michael Corsi. "Human memory and the medial temporal region of the brain." Doctoral Thesis. Doctoral Thesis. McGill University (Canada), 1972.

[196] Yihui Xie. *bookdown: Authoring books and technical documents with r markdown*. manual. 2021. URL: https://github.com/rstudio/bookdown.

[197] Max Kuhn and Davis Vaughan. *yardstick: Tidy characterizations of model performance*. manual. 2021. URL: https://CRAN.R-project.org/package=yardstick.

[198] Hadley Wickham and Jennifer Bryan. *readxl: Read excel files*. Version R package version 1.3.1. 2019. URL: https://CRAN.R-project.org/package=readxl.

[199] Xavier Robin et al. "pROC: an open-source package for R and S+ to analyze and compare ROC curves". In: *BMC bioinformatics* 12.1 (2011). Publisher: BioMed Central, pp. 1–8.

[200] Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. Vol. 1.26. tex.ids: xie_knitr_2019-1. 2019.

[201] Hadley Wickham. *ggplot2*. 2nd ed. 2016. tex.ids: wickham_ggplot2_2016-2. Cham: Springer International Publishing, 2016.

[202] Jeroen Ooms. *magick: Advanced graphics and image-processing in r*. manual. 2021. URL: https://CRAN.R-project.org/package=magick.

[203] Winston Chang. *webshot: Take screenshots of web pages*. manual. 2019. URL: https://CRAN.R-project.org/package=webshot.

[204] Hao Zhu. *kableExtra: Construct complex table with 'kable' and pipe syntax*. manual. 2021. URL: https://CRAN.R-project.org/package=kableExtra.

[205] Georgina Hobden et al. *Post-stroke executive function impairments in relation to white matter damage due to stroke lesions versus leukoaraiosis*. ISSN: 2126-6247 Type: article. medRxiv, Nov. 14, 2021, p. 2021.11.12.21266247. DOI: 10.1101/2021.11.12.21266247. URL: https://www.medrxiv.org/content/10.1101/2021.11.12.21266247v1 (visited on 02/28/2022).

[206] Teodoro del Ser et al. "Evolution of cognitive impairment after stroke and risk factors for delayed progression". In: *Stroke* 36.12 (Dec. 2005), pp. 2670–2675. DOI: 10.1161/01.STR.0000189626.71033.35.

[207] Nele Demeyere et al. "Post-stroke cognition with the Oxford Cognitive Screen vs Montreal Cognitive Assessment: a multi-site randomized controlled study (OCS-CARE)". In: *AMRC Open Research* 1.12 (2019). Publisher: F1000 Research Limited, p. 12.

[208] Milija D. Mijajlović et al. "Post-stroke dementia - a comprehensive review". In: *BMC medicine* 15.1 (Jan. 18, 2017), p. 11. DOI: 10.1186/s12916-017-0779-7.

[209] Barton W. Palmer et al. "Base Rates of "Impaired" Neuropsychological Test Performance Among Healthy Older Adults". In: *Archives of Clinical Neuropsychology* 13.6 (Aug. 1, 1998), pp. 503–511. DOI: 10.1016/S0887-6177(97)00037-1. URL: https://www.sciencedirect.com/science/article/pii/S0887617797000371 (visited on 11/07/2023).

[210] Gavin H. Taylor, Jonathan Todman, and Niall M. Broomfield. "Post-stroke emotional adjustment: A modified social cognitive transition model". In: *Neuropsychological rehabilitation* 21.6 (2011). Publisher: Taylor & Francis, pp. 808–824.

[211] Fiona J. Rowe and Lauren R. Hepworth. "The Impact of Visual Impairment in Stroke (IVIS) Study – Evidence of Reproducibility". In: *Neuro-Ophthalmology* 45.3 (), pp. 165–171. DOI: 10.1080/01658107.2020.1849317. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8210860/ (visited on 07/15/2022).

[212] Akira Miyake, Michael J. Emerson, and Naomi P. Friedman. "Assessment of executive functions in clinical settings: Problems and recommendations". In: *Seminars in speech and language*. Vol. 21. tex.ids: miyake_assessment_2000-1 section: 02. Thieme Medical Publishers, Inc, 2000, p. 169. DOI: 10.1055/s-2000-7563.

[213] H. M. Laakso et al. "Executive function subdomains are associated with post-stroke functional outcome and permanent institutionalization". In: *European Journal of Neurology* 26.3 (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ene.13854, pp. 546–552. DOI: 10.1111/ene.13854. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ene.13854 (visited on 05/18/2023).

[214] Tamara Ownsworth and David Shum. "Relationship between executive functions and productivity outcomes following stroke". In: *Disability and Rehabilitation* 30.7 (Jan. 2008), pp. 531–540. DOI: 10.1080/09638280701355694. URL: http://www.tandfonline.com/doi/full/10.1080/09638280701355694 (visited on 10/14/2020).

[215] Naomi Josman et al. "Using the virtual action planning-supermarket for evaluating executive functions in people with stroke". In: *Journal of Stroke and Cerebrovascular Diseases* 23.5 (2014). Publisher: Elsevier, pp. 879–887.

[216] Carmen S. van der Zwaluw et al. "Cognitive Functioning in the Acute Phase Poststroke: A Predictor of Discharge Destination?" In: *Journal of Stroke and Cerebrovascular Diseases* 20.6 (Nov. 1, 2011), pp. 549–555. DOI: 10.1016/j.jstrokecerebrovasdis.2010.03.009. URL: https://www.sciencedirect.com/science/article/pii/S1052305710000911 (visited on 07/18/2022).

[217] Joel J. Gagnier et al. "COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures". In: *Quality of Life Research* 30.8 (Aug. 2021), pp. 2197–2218. DOI: 10.1007/s11136-021-02822-4. URL: https://link.springer.com/10.1007/s11136-021-02822-4 (visited on 06/28/2022).

[218] Sam S. Webb and Nele Demeyere. "Using Multiverse Analysis to Highlight Differences in Convergent Correlation Outcomes Due to Data Analytical and Study Design Choices". In: *Assessment* 30.6 (Sept. 1, 2023). Publisher: SAGE Publications Inc, pp. 1825–1835. DOI: 10.1177/10731911221127904. URL: https://doi.org/10.1177/10731911221127904 (visited on 12/19/2023).

[219] Robert G. Knight et al. "Some Normative and Psychometric Data for the Geriatric Depression Scale and the Cognitive Failures Questionnaire from a Sample of Healthy Older Persons". In: *New Zealand Journal of Psychology* 33.3 (2004). Place: New Zealand Publisher: New Zealand Psychological Society, pp. 163–170.

[220] Sarah L. Green et al. "Apathy and depression as predictors of activities of daily living following stroke and traumatic brain injuries in adults: a meta-analysis". In: *Neuropsychology review* (2021). Publisher: Springer, pp. 1–19. DOI: https://doi.org/10.1007/s11065-021-09501-8.

[221] Rajesh M. Parikh et al. "The impact of poststroke depression on recovery in activities of daily living over a 2-year follow-up". In: *Archives of neurology* 47.7 (1990). Publisher: American Medical Association, pp. 785–789.

[222] Terence J. Quinn, Peter Langhorne, and David J. Stott. "Barthel Index for Stroke Trials: Development, Properties, and Application". In: *Stroke* 42.4 (Apr. 2011), pp. 1146–1151. DOI: 10.1161/STROKEAHA.110.598540. URL: https://www.ahajournals.org/doi/10.1161/STROKEAHA.110.598540 (visited on 03/18/2021).

[223] F. Mahoney and D. W. Barthel. "Functional evaluation ; the Barthel index. A simple index of the independence useful in scoring improvement in the rehabilitation of the chronically ill." In: *Maryland State Medical Journal* 14 (1965), pp. 61–65. URL: /paper/Functional-evaluation-%3B-the-Barthel-index.-A-simple-Mahoney-Barthel/c71fe40c867d7e7046e2b655cf70e12eedaac8b3 (visited on 03/16/2021).

[224] Katherine L. Salter et al. "Outcome Assessment in Randomized Controlled Trials of Stroke Rehabilitation". In: *American Journal of Physical Medicine & Rehabilitation* 86.12 (Dec. 2007), p. 1007. DOI: 10.1097/PHM.0b013e3181587b3d. URL: https://journals.lww.com/ajpmr/FullText/2007/12000/Outcome_Assessment_in_Randomized_Controlled_Trials.9.aspx?casa_token=1p6CebANOBwAAAAA:Pd17Jf0bzlGVUg4sIRUE71SbDY4aaA-R45EICkLZ357M2zIydOacA_8oQGNQpNYhYcs-N3xDfFOV_eAI3oLfSXo (visited on 10/16/2023).

[225] R.H. Harwood and S. Ebrahim. "The validity, reliability and responsiveness of the Nottingham Extended Activities of Daily Living scale in patients undergoing total hip replacement". In: *Disability and Rehabilitation* 24.7 (Jan. 2002), pp. 371–377. DOI: 10.1080/09638280110101541. URL: http://www.tandfonline.com/doi/full/10.1080/09638280110101541 (visited on 03/25/2022).

[226] Patrick Lyden. "Using the National Institutes of Health Stroke Scale". In: *Stroke* 48.2 (Feb. 2017). Publisher: American Heart Association, pp. 513–519. DOI: 10.1161/STROKEAHA.116.015434. URL: https://www.ahajournals.org/doi/10.1161/STROKEAHA.116.015434 (visited on 02/24/2022).

[227] John Rankin. "Cerebral Vascular Accidents in Patients over the Age of 60: II. Prognosis". In: *Scottish Medical Journal* 2.5 (1957). Publisher: SAGE Publications, pp. 200–215. DOI: 10.1177/003693305700200504. URL: https://doi.org/10.1177/003693305700200504 (visited on 05/06/2021).

[228] Joseph P. Broderick, Opeolu Adeoye, and Jordan Elm. "Evolution of the Modified Rankin Scale and Its Use in Future Stroke Trials". In: *Stroke* 48.7 (2017). Publisher: American Heart Association, pp. 2007–2012. DOI: 10.1161/STROKEAHA.117.017866. URL: https://www.ahajournals.org/doi/10.1161/STROKEAHA.117.017866 (visited on 05/06/2021).

[229] Abi Roper. "Computer gesture therapy for adults with severe aphasia". PhD thesis. City, University Lonon, 2017. URL: https://openaccess.city.ac.uk/id/eprint/18030/ (visited on 09/05/2022).

[230] Shelagh M. Brumfitt and Paschal Sheeran. "The development and validation of the visual analogue Self-Esteem scale (VASES) 1". In: *British Journal of Clinical Psychology* 38.4 (1999). Publisher: Wiley Online Library, pp. 387–400. DOI: https://doi.org/10.1348/014466599162980.

[231] Carmel Bradshaw, Sandra Atkinson, and Owen Doody. "Employing a Qualitative Description Approach in Health Care Research". In: *Global Qualitative Nursing Research* 4 (Nov. 24, 2017), p. 2333393617742282. DOI: 10.1177/2333393617742282. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5703087/ (visited on 10/18/2023).

[232] Meghan Ambrens et al. "Economic evaluation of the e-Health StandingTall balance exercise programme for fall prevention in people aged 70 years and over". In: *Age and Ageing* 51.6 (June 1, 2022), afac130. DOI: 10.1093/ageing/afac130. URL: https://doi.org/10.1093/ageing/afac130 (visited on 09/05/2022).

[233] Yang Tian et al. *Exploring the system-wide costs of falls in older people in Torbay.* The Kings Fund, 2013, p. 12. URL: https://www.kingsfund.org.uk/publications/exploring-system-wide-costs-falls-older-people-torbay#:~:text=While%20falls%20patients%20in%20this,local%20adult%20social%20care%20budget..

[234] J. C. Davis et al. "International comparison of cost of falls in older adults living in the community: a systematic review". In: *Osteoporosis International* 21.8 (Aug. 1, 2010), pp. 1295–1306. DOI: 10.1007/s00198-009-1162-0. URL: https://doi.org/10.1007/s00198-009-1162-0 (visited on 10/25/2022).

[235] Lakshmi Kannan et al. "Cognitive-motor exergaming for reducing fall risk in people with chronic stroke: a randomized controlled trial". In: *NeuroRehabilitation* 44.4 (2019). Publisher: IOS Press, pp. 493–510.

[236] Wycliffe E. Wei et al. "Post-stroke patients with moderate function have the greatest risk of falls: a National Cohort Study". In: *BMC Geriatrics* 19.1 (Dec. 26, 2019), p. 373. DOI: 10.1186/s12877-019-1377-7. URL: https://doi.org/10.1186/s12877-019-1377-7 (visited on 07/13/2023).

[237] Amanda J. Clark et al. "Reliability and construct validity of a revised Baycrest Multiple Errands Test". In: *Neuropsychological Rehabilitation* 27.5 (July 4, 2017), pp. 667–684. DOI: 10.1080/09602011.2015.1117981. URL: https://www.tandfonline.com/doi/full/10.1080/09602011.2015.1117981 (visited on 10/19/2021).

[238] Peter Juslin and Henrik Olsson. "Capacity Limitations and the Detection of Correlations: Comment on Kareev (2000)." In: *Psychological review* 112 (Jan. 1, 2005), 256–67, discussion 280. DOI: 10.1037/0033-295X.112.1.256.

[239] Sinéad M. Hynes et al. "Developing a computerised Multiple Elements Test for organisational difficulties". In: *International Journal of Developmental Science* 9.2 (2015). Publisher: IOS Press, pp. 85–94. URL: https://content.iospress.com/articles/international-journal-of-developmental-science/dev140157 (visited on 10/18/2023).

[240] Allen Y. Tien et al. "Computerized Wisconsin Card Sorting Test: comparison with manual administration." In: *The Kaohsiung journal of medical sciences* 12.8 (1996), pp. 479–485. URL: https://europepmc.org/article/med/8774117 (visited on 10/18/2023).

[241] Thomas D. Parsons et al. "The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology". In: *Neuropsychological rehabilitation* 27.5 (2017). Publisher: Taylor & Francis, pp. 777–807.

[242] Ralph M. Reitan and Deborah Wolfson. "Category Test and Trail Making Test as measures of frontal lobe functions". In: *The Clinical Neuropsychologist* 9.1 (1995). Publisher: Taylor & Francis, pp. 50–56.

[243] Lena Lipskaya-Velikovsky et al. "Executive functioning and daily living of individuals with chronic stroke: measurement and implications". In: *International Journal of Rehabilitation Research* 41.2 (June 2018), pp. 122–127. DOI: 10.1097/MRR.0000000000000272. URL: http://journals.lww.com/00004356-201806000-00005 (visited on 11/11/2020).

[244] Noomi Katz et al. "The Behavioural Assessment of the Dysexecutive Syndrome (BADS) in schizophrenia and its relation to functional outcomes". In: *Neuropsychological Rehabilitation* 17.2 (Apr. 2007), pp. 192–205. DOI: 10.1080/09602010600685053. URL: http://www.tandfonline.com/doi/abs/10.1080/09602010600685053 (visited on 05/11/2023).

[245] Owen A. Williams and Nele Demeyere. "Association of Depression and Anxiety With Cognitive Impairment 6 Months After Stroke". In: *Neurology* 96.15 (Apr. 13, 2021), e1966–e1974. DOI: 10.1212/WNL.0000000000011748.

[246] Pamela Enderby, Alexandra John, and Brian Petheram. *Therapy outcome measures for rehabilitation professionals: speech and language therapy, physiotherapy, occupational therapy*. John Wiley & Sons, 2013.

[247] F. M. Nouri and N. B. Lincoln. "An extended activities of daily living scale for stroke patients". In: *Clinical rehabilitation* 1.4 (1987). Publisher: Sage Publications Sage CA: Thousand Oaks, CA, pp. 301–305.

[248] Hadley Wickham et al. "Welcome to the Tidyverse". In: *Journal of Open Source Software* 4.43 (Nov. 21, 2019), p. 1686. DOI: 10.21105/joss.01686. URL: https://joss.theoj.org/papers/10.21105/joss.01686 (visited on 03/18/2021).

[249] Claus O. Wilke. *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. Version R package version 1.1.1. 2020. URL: https://CRAN.R-project.org/package=cowplot.

[250] Hadley Wickham. *httr: Tools for Working with URLs and HTTP*. Version R package version 1.4.2. 2022. URL: https://CRAN.R-project.org/package=httr.

[251] Constantin Ahlmann-Eltze and Indrajeet Patil. *ggsignif: Significance Brackets for 'ggplot2'*. manual. 2021. URL: https://CRAN.R-project.org/package=ggsignif.

[252] Frank E. Harrell Jr. *Hmisc: Harrell Miscellaneous*. Version R package version 4.5.0. 2021. URL: https://CRAN.R-project.org/package=Hmisc.

[253] Matt Dowle and Arun Srinivasan. *data.table: Extension of 'data.frame'*. manual. 2021. URL: https://CRAN.R-project.org/package=data.table.

[254] Baptiste Auguie. *gridExtra: Miscellaneous functions for "Grid" graphics*. manual. 2017. URL: https://CRAN.R-project.org/package=gridExtra.

[255] Tamar Abzhandadze et al. "Very Early MoCA Can Predict Functional Dependence at 3 Months After Stroke: A Longitudinal, Cohort Study". In: *Frontiers in Neurology* 10 (2019). URL: https://www.frontiersin.org/articles/10.3389/fneur.2019.01051 (visited on 07/18/2023).

[256] Vera Zietemann et al. "Early MoCA predicts long-term cognitive and functional outcome and mortality after stroke". In: *Neurology* 91.20 (Nov. 13, 2018), e1838–e1850. DOI: 10.1212/WNL.0000000000006506. URL: https://www.neurology.org/lookup/doi/10.1212/WNL.0000000000006506 (visited on 07/18/2023).

[257] M. Tracy Morrison et al. "Multiple Errands Test–Revised (MET–R): A performance-based measure of executive function in people with mild cerebrovascular accident". In: *American Journal of Occupational Therapy* 67.4 (2013). Publisher: American Occupational Therapy Association, pp. 460–468.

[258] Adina Maeir, Shira Krauss, and Noomi Katz. "Ecological Validity of the Multiple Errands Test (MET) on Discharge from Neurorehabilitation Hospital". In: *OTJR: Occupation, Participation and Health* 31.1 (Jan. 1, 2011). Publisher: SAGE Publications Inc, S38–S46. DOI: 10.3928/15394492-20101108-07. URL: https://doi.org/10.3928/15394492-20101108-07 (visited on 10/19/2021).

[259] Katie Spencer et al. "The impact of the COVID-19 pandemic on radiotherapy services in England, UK: a population-based study". In: *The Lancet Oncology* 22.3 (Mar. 1, 2021), pp. 309–320. DOI: 10.1016/S1470-2045(20)30743-9. URL: https://www.sciencedirect.com/science/article/pii/S1470204520307439 (visited on 01/25/2024).

[260] Nishita Padmanabhan et al. "Impact of COVID-19 on stroke admissions, treatments, and outcomes at a comprehensive stroke centre in the United Kingdom". In: *Neurological Sciences* 42.1 (Jan. 1, 2021), pp. 15–20. DOI: 10.1007/s10072-020-04775-x. URL: https://doi.org/10.1007/s10072-020-04775-x (visited on 01/25/2024).

[261] Aitor Alquézar-Arbé et al. "Impact of the COVID-19 pandemic on hospital emergency departments: results of a survey of departments in 2020 - the Spanish ENCOVUR study." In: *Emergencias* (2020), pp. 320–331. URL: https://pesquisa.bvsalud.org/gim/resource/en/mdl-33006832?src=similardocs (visited on 11/17/2021).

[262] Abdel Douiri et al. "Stroke Care in the United Kingdom During the COVID-19 Pandemic". In: *Stroke* 52.6 (June 2021). Publisher: American Heart Association, pp. 2125–2133. DOI: 10.1161/STROKEAHA.120.032253. URL: https://www.ahajournals.org/doi/full/10.1161/STROKEAHA.120.032253 (visited on 03/31/2022).

[263] A. Bersano et al. "Stroke care during the COVID-19 pandemic: experience from three large European countries". In: *European Journal of Neurology* 27.9 (2020). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ene.14375, pp. 1794–1800. DOI: 10.1111/ene.14375. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ene.14375 (visited on 01/25/2024).

[264] Melissa E. Parlar et al. ""You can't touch this": Delivery of inpatient neuropsychological assessment in the era of COVID-19 and beyond". In: *The Clinical Neuropsychologist* 34.7 (Nov. 16, 2020), pp. 1395–1410. DOI: 10.1080/13854046.2020.1810324. URL: https://www.tandfonline.com/doi/full/10.1080/13854046.2020.1810324 (visited on 01/25/2024).

[265] C Montani et al. "Psychological impact of a remote psychometric consultation with hospitalized elderly people". In: *Journal of Telemedicine and Telecare* 3.3 (Sept. 1, 1997). Publisher: SAGE Publications, pp. 140–145. DOI: 10.1258/1357633971931048. URL: https://doi.org/10.1258/1357633971931048 (visited on 01/25/2024).

[266] Sam Parsons, Anne-Wil Kruijt, and Elaine Fox. "Psychological Science Needs a Standard Practice of Reporting the Reliability of Cognitive-Behavioral Measurements". In: *Advances in Methods and Practices in Psychological Science* 2.4 (Dec. 1, 2019). Publisher: SAGE Publications Inc, pp. 378–395. DOI: 10.1177/2515245919879695. URL: https://doi.org/10.1177/2515245919879695 (visited on 09/15/2021).

[267] Ching-Lan Chin and Grace Yao. "Convergent validity". In: ed. by Alex C. Michalos. Encyclopedia of quality of life and well-being research. Dordrecht: Springer Netherlands, 2014, pp. 1275–1276. DOI: 10.1007/978-94-007-0753-5_573. URL: https://doi.org/10.1007/978-94-007-0753-5_573.

[268] Frank Ho-yin Lai et al. "The validity, reliability and clinical utility of a performance-based executive function assessment in people with mild to moderate dementia". In: *Aging & Mental Health* 24.9 (Sept. 1, 2020), pp. 1496–1504. DOI: 10.1080/13607863.2019.1599818. URL: https://www.tandfonline.com/doi/full/10.1080/13607863.2019.1599818 (visited on 08/09/2023).

[269] Katri Saar et al. "Validation of a New Cognitive Screening Method for Stroke Patients". In: *Behavioural Neurology* 2019 (Nov. 6, 2019), p. 2943603. DOI: 10.1155/2019/2943603. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6875250/ (visited on 02/28/2022).

[270] Sam Sappho Webb and Nele Demeyere. "Comparing the Oxford Digital Multiple Errands Test (OxMET) to a real-life version: convergence, feasibility, and acceptability." In: (Dec. 23, 2023). DOI: 10.31234/osf.io/4c5et.

[271] Sam Sappho Webb and Nele Demeyere. *Predictive validity of the Oxford Digital Multiple Errands Test (OxMET) for functional outcomes after stroke*. Feb. 16, 2023. DOI: 10.31234/osf.io/agb3j. URL: https://psyarxiv.com/agb3j/ (visited on 08/01/2023).

[272] E Kaplan, H Goodglass, and S Weintraud. *Boston Naming Test*. Philadelphia: Lee and Febiger, 1983.

[273] Brian P. Yochim, Katherine D. Kane, and Anne E. Mueller. "Naming Test of the Neuropsychological Assessment Battery: Convergent and Discriminant Validity". In: *Archives of Clinical Neuropsychology* 24.6 (Sept. 1, 2009), pp. 575–583. DOI: 10.1093/arclin/acp053. URL: https://doi.org/10.1093/arclin/acp053 (visited on 01/25/2024).

[274] Alexandra Harry and Simon F. Crowe. "Is the Boston Naming Test Still Fit For Purpose?" In: *The Clinical Neuropsychologist* 28.3 (Apr. 3, 2014), pp. 486–504. DOI: 10.1080/13854046.2014.892155. URL: http://www.tandfonline.com/doi/abs/10.1080/13854046.2014.892155 (visited on 01/25/2024).

[275] Benjamin Lam et al. "Criterion and Convergent Validity of the Montreal Cognitive Assessment with Screening and Standardized Neuropsychological Testing". In: *Journal of the American Geriatrics Society* 61.12 (2013). _eprint: https://agsjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/jgs.12541, pp. 2181–2185. DOI: 10.1111/jgs.12541. URL: https://agsjournals.onlinelibrary.wiley.com/doi/abs/10.1111/jgs.12541 (visited on 09/06/2021).

[276] Anne M. Damian et al. "The Montreal Cognitive Assessment and the Mini-Mental State Examination as Screening Instruments for Cognitive Impairment: Item Analyses and Threshold Scores". In: *Dementia and Geriatric Cognitive Disorders* 31.2 (2011), pp. 126–131. DOI: 10.1159/000323867. URL: https://www.karger.com/Article/FullText/323867 (visited on 01/28/2024).

[277] Brenda Hanae Bentes Koshimoto et al. "Floor and ceiling effects on the Montreal Cognitive Assessment in patients with Parkinson's disease in Brazil". In: *Dementia & Neuropsychologia* 17 (2023). Publisher: SciELO Brasil, e20230022. URL: https://www.scielo.br/j/dn/a/GrcWSQxYThdPZtTjPLWWJbK/ (visited on 01/28/2024).

[278] S.T. Pendlebury et al. "Differences in Cognitive Profile between TIA, Stroke and Elderly Memory Research Subjects: A Comparison of the MMSE and MoCA". In: *Cerebrovascular Diseases* 34.1 (2012), pp. 48–54. DOI: 10.1159/000338905. URL: https://www.karger.com/Article/FullText/338905 (visited on 01/28/2024).

[279] R. Christopher Branson. "Examining the relationship among measures of global cognition, executive function, and instrumental activities of daily living: can they all just get along?" Publisher: University of Tennessee at Chattanooga. PhD thesis. Chattanooga (Tenn.): University of Tennessee at Chattanooga, 2016. URL: https://scholar.utc.edu/theses/447/.

[280] Krystyna Jaracz et al. "Post-stroke quality of life and depression". In: *Acta Neuropsychiatrica* 14.5 (2002). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1034/j.1601-5215.2002.140504.x, pp. 219–225. DOI: 10.1034/j.1601-5215.2002.140504.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1034/j.1601-5215.2002.140504.x (visited on 01/26/2024).

[281] Anners Lerdal and Caryl L. Gay. "Acute-phase fatigue predicts limitations with activities of daily living 18 months after first-ever stroke". In: *Journal of Stroke and Cerebrovascular Diseases* 26.3 (2017). Publisher: Elsevier, pp. 523–531. URL: https://www.sciencedirect.com/science/article/pii/S1052305716305961?casa_token=-r-zF0WATswAAAAA:hozbVFG2c7KTUbIPmtcFHcVOMmiNPWpEE8g8MpEPK7bpbDXTxVfSH5ed4WotyYZ8wtKWrftM (visited on 01/26/2024).

[282] Stanley L. Brodsky and Bronwen Lichtenstein. "The Gold Standard and the Pyrite Principle: Toward a Supplemental Frame of Reference". In: *Frontiers in Psychology* 11 (Mar. 31, 2020), p. 562. DOI: 10.3389/fpsyg.2020.00562. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7136847/ (visited on 01/26/2024).

[283] Scott O. Lilienfeld et al. "Fifty psychological and psychiatric terms to avoid: a list of inaccurate, misleading, misused, ambiguous, and logically confused words and phrases". In: *Frontiers in Psychology* 6 (2015). URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01100 (visited on 01/26/2024).

[284] Lee J. Cronbach and Paul E. Meehl. "Construct validity in psychological tests". In: *Psychological Bulletin* 52.4 (1955). Place: US Publisher: American Psychological Association, pp. 281–302. DOI: 10.1037/h0040957.

[285] Sally J. Vogel et al. "Concordance of the Montreal cognitive assessment with standard neuropsychological measures". In: *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 1.3 (2015). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1016/j.dadm.2015.05.002, pp. 289–294. DOI: 10.1016/j.dadm.2015.05.002. URL: https://onlinelibrary.wiley.com/doi/abs/10.1016/j.dadm.2015.05.002 (visited on 01/26/2024).

[286] Tresa M. Roebuck-Spencer et al. "Cognitive Screening Tests Versus Comprehensive Neuropsychological Test Batteries: A National Academy of Neuropsychology Education Paper†". In: *Archives of Clinical Neuropsychology* 32.4 (June 1, 2017), pp. 491–498. DOI: 10.1093/arclin/acx021. URL: https://doi.org/10.1093/arclin/acx021 (visited on 01/26/2024).

[287] Naomi H. Martin et al. "The neuropsychology needs of a hyper-acute stroke unit". In: *Journal of the Neurological Sciences* 423 (Apr. 15, 2021), p. 117382. DOI: 10.1016/j.jns.2021.117382. URL: https://www.sciencedirect.com/science/article/pii/S0022510X21000757 (visited on 01/28/2024).

[288] M. J. E. van Zandvoort et al. "Early neuropsychological evaluation in patients with ischaemic stroke provides valid information". In: *Clinical Neurology and Neurosurgery* 107.5 (Aug. 1, 2005), pp. 385–392. DOI: 10.1016/j.clineuro.2004.10.012. URL: https://www.sciencedirect.com/science/article/pii/S0303846704001659 (visited on 01/28/2024).

[289] Joseph D. Matarazzo. “Matarazzo, Joseph D. "Psychological assessment versus psychological testing: Validation from Binet to the school, clinic, and courtroom." In: *American Psychologist* 45.9 (1990), p. 999. URL: https://scholar.google.com/scholar_lookup?title=Psychological%20assessment%20versus%20psychological%20testing%3A%20Validation%20from%20Binet%20to%20the%20school%2C%20clinic%20and%20courtroom&author=J.%20Matarazzo&publication_year=1990&journal=American%20Psychologist&volume=45&pages=999-1017 (visited on 01/26/2024).

[290] Robert A. Stern and Travis White. *NAB, Neuropsychological Assessment Battery: Administration, scoring, and interpretation manual*. Lutz: Psychological Assessment Resources Odessa, FL.

[291] Psychological Assessment Resources. *Neuropsychological Assessment Battery | NAB*. parinc. 2024. URL: https://www.parinc.com/Products?pkey=260 (visited on 01/26/2024).

[292] Pearson. *Behavioural Assessment of the Dysexecutive Syndrome*. pearsonclinical. 2024. URL: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Behavioural-Assessment-of-the-Dysexecutive-Syndrome/p/P100009053.html (visited on 01/26/2024).

[293] Pearson. *Repeatable Battery for the Assessment of Neuropsychological Status Update*. pearsonclinical. 2024. URL: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Repeatable-Battery-for-the-Assessment-of-Neuropsychological-Status-Update/p/P100009272.html (visited on 01/26/2024).

[294] Vladimir Hachinski et al. “National Institute of Neurological Disorders and Stroke–Canadian Stroke Network Vascular Cognitive Impairment Harmonization Standards”. In: *Stroke* 37.9 (Sept. 2006). Publisher: American Heart Association, pp. 2220–2241. DOI: 10.1161/01.STR.0000237236.88823.47. URL: https://www.ahajournals.org/doi/full/10.1161/01.str.0000237236.88823.47 (visited on 01/28/2024).

[295] Donald T Campbell and Donald W Fiske. “Convergent and discriminant validation by the multitrait-multimethod matrix”. In: *Psychological Bulletin* 56.2 (1959).

[296] Robert P. Althauser and Thomas A. Heberlein. “Validity and the multitrait-multimethod matrix”. In: *Sociological methodology* 2 (1970). Publisher: JSTOR, pp. 151–169. URL: https://www.jstor.org/stable/270788?casa_token=a3LQE_6756cAAAAA:dUAAvymvU3SLFJwolj2ReaftO8iQASJuntyNJWz7oglICzP1oQXYWVzCKnaddjXRoqUdw19m8lQ9kAENzZIafsOyBYevLk1SE_H5Q (visited on 01/26/2024).

[297] Arne L. Kalleberg and James R. Kluegel. “Analysis of the multitrait-multimethod matrix: Some limitations and an alternative.” In: *Journal of Applied Psychology* 60.1 (1975). Publisher: American Psychological Association, p. 1. URL: https://psycnet.apa.org/journals/apl/60/1/1/ (visited on 01/26/2024).

[298] Chia-Yeh Chou et al. "Comparison of construct validity of two short forms of Stroke-Specific Quality of Life scale". In: *PLOS ONE* 12.12 (Dec. 6, 2017). Ed. by Mohd Noor Norhayati, e0188478. DOI: 10.1371/journal.pone.0188478. URL: https://dx.plos.org/10.1371/journal.pone.0188478 (visited on 01/26/2024).

[299] Andreea M. Rawlings et al. "Factor Structure of the ARIC-NCS Neuropsychological Battery: An Evaluation of Invariance Across Vascular Factors and Demographic Characteristics". In: *Psychological assessment* 28.12 (Dec. 2016), pp. 1674–1683. DOI: 10.1037/pas0000293. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018233/ (visited on 01/26/2024).

[300] D Johnson et al. "Reliability and construct validity of the Automated Neuropsychological Assessment Metrics (ANAM) mood scale". In: *Archives of Clinical Neuropsychology* 23.1 (Jan. 2008), pp. 73–85. DOI: 10.1016/j.acn.2007.10.001. URL: https://academic.oup.com/acn/article-lookup/doi/10.1016/j.acn.2007.10.001 (visited on 01/26/2024).

[301] Allan F. Mirsky et al. "Analysis of the elements of attention: A neuropsychological approach". In: *Neuropsychology review* 2 (1991). Publisher: Springer, pp. 109–145. URL: https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/BF01109051&casa_token=akHDGzJDd8AAAAAA:ad_X3vrP19A2q0o1nUxSQDL0zBOpLupJlkAO44uG7Nj_UIuQXRXArG4xzYJX1yVmOqKesNKt9tV7_dc (visited on 01/26/2024).

[302] Noelle E. Carlozzi et al. "Factor Analysis of the Repeatable Battery for the Assessment of Neuropsychological Status". In: *Applied Neuropsychology* 15.4 (Dec. 3, 2008), pp. 274–279. DOI: 10.1080/09084280802325124. URL: http://www.tandfonline.com/doi/abs/10.1080/09084280802325124 (visited on 01/26/2024).

[303] Kevin Duff et al. "Examining the repeatable battery for the assessment of neuropsychological status: Factor analytic studies in an elderly sample". In: *The American Journal of Geriatric Psychiatry* 14.11 (2006). Publisher: Elsevier, pp. 976–979. URL: https://www.sciencedirect.com/science/article/pii/S106474811260877X?casa_token=UzP3IRUskMgAAAAA:AOtnaAUUGhLWDTQSPN1csdZ06Snog7770-Bm-aJrwYSC8FoW_SeBpntV9-pHx1R9BeRU8jLq (visited on 01/26/2024).

[304] M. C. Wilde. "The Validity of the Repeatable Battery of Neuropsychological Status in Acute Stroke". In: *The Clinical Neuropsychologist* 20.4 (Dec. 2006), pp. 702–715. DOI: 10.1080/13854040500246901. URL: http://www.tandfonline.com/doi/abs/10.1080/13854040500246901 (visited on 01/26/2024).

[305] Ercole Vellone et al. "Psychometric evaluation of the stroke impact scale 3.0". In: *Journal of Cardiovascular Nursing* 30.3 (2015). Publisher: LWW, pp. 229–241. URL: https://journals.lww.com/jcnjournal/FullText/2015/05000/Psychometric_Evaluation_of_the_Stroke_Impact_Scale.12.aspx?casa_token=_oNn3dfGSZcAAAAA:Jg5m3vHHEKiOqPiSpd2lcZf1px1IlmYa_eFfse2KMmyMFTDYgXgIQk3OTaGAK01lvDRsLVMG9CTCDZF-o2lFOw&casa_token=awTRlScjr7MAAAAA:IzLI09KNdKw5usCCamm2Inp_

x1mL567633IR41GKwC6DFgJlUeR1qdPsI481uoAMe_P_u-NQZj3t4GsSiRW4_Q (visited on 01/26/2024).

[306] John E. Ware Jr and Cathy Donald Sherbourne. "The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection". In: *Medical care* (1992). Publisher: JSTOR, pp. 473–483. URL: https://www.jstor.org/stable/3765916?casa_token=sBIBqZteO48AAAAA:bQlGPvEK1ujZr3D-J9TtUp2i1H-Ndx9XpkpkDHGdAWjZUagXff7EnuI2QL5cKd65VQvw6pC5FOWfuM4WGKOAZeEIZDGPVIs2kATnafaPInJ78d (visited on 01/26/2024).

[307] A. S. Zigmond and R. P. Snaith. "The Hospital Anxiety and Depression Scale". In: *Acta Psychiatrica Scandinavica* 67.6 (June 1983), pp. 361–370. DOI: 10.1111/j.1600-0447.1983.tb09716.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1600-0447.1983.tb09716.x (visited on 01/24/2024).

[308] Timothy A. Brown and Michael T. Moore. "Confirmatory factor analysis". In: *Handbook of structural equation modeling* 361 (2012). Publisher: Guilford, p. 379. URL: https://books.google.co.uk/books?hl=en&lr=&id=P16bEAAAQBAJ&oi=fnd&pg=PA261&dq=confirmatory+factor+analysis+in+psychometrics&ots=jz9oAmbw6N&sig=soIQ_Yj8ZuvhaX1h7yZ5YZ1WBtI (visited on 01/26/2024).

[309] Urbano Lorenzo-Seva. "SOLOMON: a method for splitting a sample into equivalent subsamples in factor analysis". In: *Behavior Research Methods* 54.6 (Dec. 1, 2022), pp. 2665–2677. DOI: 10.3758/s13428-021-01750-y. URL: https://doi.org/10.3758/s13428-021-01750-y (visited on 01/26/2024).

[310] Marjolein Fokkema and Samuel Greiff. "How Performing PCA and CFA on the Same Data Equals Trouble". In: *European Journal of Psychological Assessment* 33.6 (Nov. 2017). Publisher: Hogrefe Publishing, pp. 399–402. DOI: 10.1027/1015-5759/a000460. URL: https://econtent.hogrefe.com/doi/10.1027/1015-5759/a000460 (visited on 01/26/2024).

[311] John Bamford et al. "Classification and natural history of clinically identifiable subtypes of cerebral infarction". In: *The Lancet* 337.8756 (1991), pp. 1521–1526. DOI: https://doi.org/10.1016/0140-6736(91)93206-O.

[312] National Institute for Health and Care Excellence. *What is the prevalence of stroke and TIA in the UK?* 2023. URL: https://cks.nice.org.uk/topics/stroke-tia/background-information/prevalence/ (visited on 01/26/2024).

[313] Office for National Statistics (ONS). *Ethnic group, England and Wales: Census 2021*. Nov. 29, 2022. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021#cite-this-statistical-bulletin.

[314] Office for National Statistics (ONS). *Profile of the older population living in England and Wales in 2021 and changes since 2011 - Office for National Statistics*. census2021. 2021. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/

profileoftheolderpopulationlivinginenglandandwalesin2021andchangessince2011/2023-04-03 (visited on 01/28/2024).

[315] Bruno Bonnechère et al. "Does ethnicity influence dementia, stroke and mortality risk? Evidence from the UK Biobank". In: *Frontiers in Public Health* 11 (2023). URL: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1111321 (visited on 01/27/2024).

[316] Amy E. Maxwell et al. "Reasons for non-recruitment of eligible patients to a randomised controlled trial of secondary prevention after intracerebral haemorrhage: observational study". In: *Trials* 18.1 (Apr. 5, 2017), p. 162. DOI: 10.1186/s13063-017-1909-4. URL: https://doi.org/10.1186/s13063-017-1909-4 (visited on 01/27/2024).

[317] Matthew Anthony Kirkman et al. "Difficulties with recruiting into neurosurgical clinical trials: The Surgical Trial in IntraCerebral Haemorrhage II as an example". In: *British Journal of Neurosurgery* 25.2 (Apr. 2011), pp. 231–234. DOI: 10.3109/02688697.2010.539718. URL: http://www.tandfonline.com/doi/full/10.3109/02688697.2010.539718 (visited on 01/27/2024).

[318] Klaus Kaae Andersen et al. "Hemorrhagic and Ischemic Strokes Compared". In: *Stroke* 40.6 (June 2009). Publisher: American Heart Association, pp. 2068–2072. DOI: 10.1161/STROKEAHA.108.540112. URL: https://www.ahajournals.org/doi/full/10.1161/STROKEAHA.108.540112 (visited on 01/27/2024).

[319] Sarah T. Pendlebury, Peter M. Rothwell, and Oxford Vascular Study. "Incidence and prevalence of dementia associated with transient ischaemic attack and stroke: analysis of the population-based Oxford Vascular Study". In: *The Lancet. Neurology* 18.3 (Mar. 2019), pp. 248–258. DOI: 10.1016/S1474-4422(18)30442-3.

[320] Richa Sharma et al. "Early Post-stroke Cognition: In-hospital Predictors and the Association With Functional Outcome". In: *Frontiers in Neurology* 11 (2020). URL: https://www.frontiersin.org/articles/10.3389/fneur.2020.613607 (visited on 01/28/2024).

[321] Satoshi Suda et al. "Early Cognitive Impairment after Minor Stroke: Associated Factors and Functional Outcome". In: *Journal of Stroke and Cerebrovascular Diseases* 29.5 (May 1, 2020), p. 104749. DOI: 10.1016/j.jstrokecerebrovasdis.2020.104749. URL: https://www.sciencedirect.com/science/article/pii/S1052305720301269 (visited on 01/28/2024).

[322] Myzoon Ali et al. "Representation of People with Aphasia in Randomized Controlled Trials of Acute Stroke Interventions". In: *International Journal of Stroke* 9.2 (Feb. 1, 2014). Publisher: SAGE Publications, pp. 174–182. DOI: 10.1111/ijs.12043. URL: https://doi.org/10.1111/ijs.12043 (visited on 01/27/2024).

[323] Benjamin Hotter et al. "Selection bias in clinical stroke trials depending on ability to consent". In: *BMC Neurology* 17.1 (Dec. 2017), p. 206. DOI: 10.1186/s12883-017-0989-9. URL: https://bmcneurol.biomedcentral.com/articles/10.1186/s12883-017-0989-9 (visited on 01/27/2024).

[324] Paula M. Janssen et al. "Neurological Deficits in Stroke Patients that May Impede the Capacity to Provide Informed Consent for Endovascular Treatment Trials". In: *Journal of Stroke and Cerebrovascular Diseases* 28.12 (Dec. 1, 2019), p. 104447. DOI: 10.1016/j.jstrokecerebrovasdis.2019.104447. URL: https://www.sciencedirect.com/science/article/pii/S1052305719305294 (visited on 01/27/2024).

[325] Nayaar Islam et al. "Accuracy of the Montreal Cognitive Assessment tool for detecting mild cognitive impairment: A systematic review and meta-analysis". In: *Alzheimer's & Dementia* 19.7 (2023). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/alz.13040, pp. 3235–3243. DOI: 10.1002/alz.13040. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.13040 (visited on 01/27/2024).

[326] Dan Shi, Xiao Chen, and Zheng Li. "Diagnostic test accuracy of the Montreal Cognitive Assessment in the detection of post-stroke cognitive impairment under different stages and cutoffs: a systematic review and meta-analysis". In: *Neurological Sciences* 39.4 (Apr. 1, 2018), pp. 705–716. DOI: 10.1007/s10072-018-3254-0. URL: https://doi.org/10.1007/s10072-018-3254-0 (visited on 01/27/2024).

[327] Marco Iosa et al. "Principal Component Analysis of Oxford Cognitive Screen in Patients With Stroke". In: *Frontiers in neurology* 13 (2022). Publisher: Frontiers Media SA.

[328] Frances Chumney. *Principal components analysis, exploratory factor analysis, and confirmatory factor analysis*. 2012. URL: https://www.westga.edu/academics/research/vrc/assets/docs/PCA-EFA-CFA_EssayChumney_09282012.pdf (visited on 01/26/2024).

[329] Svante Wold, Kim Esbensen, and Paul Geladi. "Principal component analysis". In: *Chemometrics and intelligent laboratory systems* 2.1 (1987). Publisher: Elsevier, pp. 37–52. URL: https://www.sciencedirect.com/science/article/pii/0169743987800849 (visited on 01/26/2024).

[330] Antonio Luigi Bisogno et al. "A low-dimensional structure of neurological impairment in stroke". In: *Brain Communications* 3.2 (Apr. 5, 2021), fcab119. DOI: 10.1093/braincomms/fcab119. URL: https://academic.oup.com/braincomms/article/doi/10.1093/braincomms/fcab119/6291650 (visited on 01/26/2024).

[331] Thomas Brott et al. "Measurements of acute cerebral infarction: a clinical examination scale." In: *Stroke* 20.7 (1989). Publisher: Am Heart Assoc, pp. 864–870. DOI: https://doi.org/10.1161/01.STR.20.7.864.

[332] Donald A. Norman and Tim Shallice. "Attention to action". In: Consciousness and self-regulation. tex.ids: norman_attention_1986-1. Springer, 1986, pp. 1–18.

[333] Tom Steverson. "Ecological assessment of the supervisory attentional system in people with intellectual disabilities." PhD thesis. University of East Anglia, 2011. URL: https://ueaeprints.uea.ac.uk/id/eprint/39454/1/2011SteversonTDClinPsyc.pdf (visited on 01/26/2024).

[334] John Brooke. "SUS: A 'Quick and Dirty' Usability Scale". In: *Usability Evaluation In Industry*. 1st ed. London: CRC Press, 1996. URL: https://doi.org/10.1201/9781498710411.

[335] Donnchadh Murphy et al. "Oxford cognitive screen: A critical review and independent psychometric evaluation". In: *Journal of Neuropsychology* 17.3 (2023). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jnp.12318, pp. 491–504. DOI: 10.1111/jnp.12318. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jnp.12318 (visited on 01/27/2024).

[336] Wai-Ling Bickerton et al. "The BCoS cognitive profile screen: Utility and predictive value for stroke." In: *Neuropsychology* 29.4 (July 2015), pp. 638–648. DOI: 10.1037/neu0000160. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/neu0000160 (visited on 10/10/2023).

[337] Robert Katzman et al. "Validation of a short Orientation-Memory-Concentration Test of cognitive impairment." In: *The American journal of psychiatry* 140.6 (1983), pp. 734–739.

[338] Carolyn M. Baum et al. "Reliability, validity, and clinical utility of the Executive Function Performance Test: A measure of executive function in a sample of people with stroke". In: *The American journal of occupational therapy* 62.4 (2008). Publisher: The American Occupational Therapy Association, Inc., pp. 446–455.

[339] A. Smith. "The Digit Modalities Test: a new psychologic test for economic screening of learning and other cerebral disorders". In: *Learning disorders* 3 (1968), pp. 83–91.

[340] M. Tracy Morrison, Dorothy F. Edwards, and Gordon Muir Giles. "Performance-based testing in mild stroke: Identification of unmet opportunity for occupational therapy". In: *The American Journal of Occupational Therapy* 69.1 (2015). Publisher: The American Occupational Therapy Association, Inc., 6901360010p1–6901360010p5.

[341] Birk Diedenhofen and Jochen Musch. "cocron: A Web Interface and R Package for the Statistical Comparison of Cronbach's Alpha Coefficients." In: *International Journal of Internet Science* 11.1 (2016).

[342] Emily J. Nalder et al. "Clinicians' perceptions of the clinical utility of the Multiple Errands Test for adults with neurological conditions". In: *Neuropsychological Rehabilitation* 27.5 (July 4, 2017). Publisher: Routledge _eprint: https://doi.org/10.1080/09602011.2015.1067628, pp. 685–706. DOI: 10.1080/09602011.2015.1067628. URL: https://doi.org/10.1080/09602011.2015.1067628 (visited on 11/11/2020).

[343] Shannon M. Scarff et al. “The Multiple Errands Test: A Guide for Site-Specific Version Development”. In: *Canadian Journal of Occupational Therapy* (Nov. 29, 2022). Publisher: SAGE Publications Inc, p. 00084174221142184. DOI: 10.1177/00084174221142184. URL: https://doi.org/10.1177/00084174221142184 (visited on 12/13/2022).