The Oxford Cognitive Screen – Plus (OCS-Plus): a tablet based short cognitive screening tool for milder cognitive impairment

Nele Demeyere[1], Marleen Haupt[2], Sam Webb[1], Lea Strobel[2], Elise Milosevich[1], Margaret J Moore[1], Hayley Wright[3], Kathrin Finke[4], Mihaela Duta[1]

[1] *Department of Experimental Psychology, University of Oxford*

[2] *Department of General and Experimental Psychology, Ludwig-Maximilians-Universität München, Munich, Germany*

[3] *Coventry University*

[4] *Hans-Berger Department of Neurology, University Hospital Jena, Jena, Germany*

Correspondence to:

Nele Demeyere

Department of Experimental Psychology

New Radcliffe House, Radcliffe Observatory Quarter

Oxford, OX2 6AE

Tel: 01865 271340

Email: nele.demeyere@psy.ox.ac.uk

Acknowledgements

This work was funded by a Stroke Association UK award to ND (TSA2015_LECT02 and SA PPA 18\100032) and was supported by the National Institute for Health Research (NIHR) Oxford Health Biomedical Research Centre (BRC) at Oxford University.

We would like to express our sincere gratitude and admiration to the late Prof Glyn W Humphreys, who initiated the OCS-Plus work. We also thank the contributors to the data collection both in Munich, Coventry and Oxford.

The Oxford Cognitive Screen – Plus (OCS-Plus): a tablet based short cognitive screening tool for milder cognitive impairment.

Abstract

The Oxford Cognitive Screen was developed as a brief screening tool for common post-stroke focal cognitive deficits, including language, memory, attention, praxis, and number-processing impairments. Here, we present the OCS-Plus, a computerised tablet-based screen designed to briefly assess domain-general cognition and provide more fine-grained measures of memory and executive function. The OCS-Plus was designed to sensitively screen for subtle cognitive impairments and differentiate between memory and non-memory deficits.

The OCS-Plus contains 10 subtasks and requires approximately 30 minutes to complete. In this study, 320 neurologically healthy ageing participants (age *M*=62.66*, SD*=13.75) from three sites completed the OCS-Plus to provide a normative sample. The convergent validity of this assessment was established in comparison to the ACE-R, CERAD and Rey-Osterrieth. Divergent validity was established through comparison with the BDI. Internal consistency of each subtask was evaluated and test-retest reliability was determined.

We established the normative impairment cut-offs for each of the subtests. Predicted convergent and divergent validity was found as well as strong test-retest reliability, which provided evidence of test stability. Further research demonstrating the use and validity of the OCS-Plus in various clinical populations is required.

The OCS-Plus is a standardised cognitive screening tool, normed and validated in a large sample of neurologically healthy participants. The OCS-Plus is available as an Android App and provides an automated report of domain-general cognitive impairments in executive attention and memory.

Keywords: Cognition, mild cognitive impairment, cognitive screening, tablet-based screening

The Oxford Cognitive Screen – Plus (OCS-Plus): a tablet based short cognitive screening tool for mild cognitive impairment

As individuals age, they undergo a series of biological and psychological changes. To some degree, this natural development leads to changes in brain structure as well as function and, thereby, also in cognition. However, the impact of these age-related changes is not uniform across all brain regions, cognitive domains, or individuals (Glisky, 2007).

While cognitive decline is a natural part of the ageing process, a subset of ageing adults exhibit an abnormally severe decline in cognitive abilities resulting in cognitive impairment. Cognitive impairment is defined as demonstrable impaired cognitive abilities surpassing what is considered 'normal' within the ageing process (Mosti et al., 2019). Cognitive impairment is a spectrum ranging from very mild cognitive deficits to severe, debilitating dementia (Dichgans, 2017). Subtle cognitive impairment can take the form of mild, yet evident reductions in episodic memory, processing speed, working memory capacity, and/or verbal fluency (Finkel, Reynolds, McArdle, & Pederson, 2007; Mosti, Rog, & Fink, 2019). These subtle cognitive impairments are generally not a stable phenomenon, but can fluctuate in severity over time (Dichgans, 2017), which makes identifying these subtle cognitive deficits particularly challenging. However, an early diagnosis of subtle cognitive impairments is essential as it can considerably impact early guidance of the patients and relatives, as well as facilitate the identification of comorbidities or potential alternative explanations causing cognitive decline. Importantly, some patient populations are characterized by a higher risk to develop dementia. Stroke survivors commonly exhibit cognitive deficits and have a higher probability of developing dementia than age-matched individuals with no history of stroke (Busse, Angermeyer, & Riedel-Heller, 2006; Merriman et al., 2019).

Similarly, patients with Mild Cognitive Impairment (MCI) - defined by cognitive impairments exceeding age norms but not interfering with an independent everyday life

(Alzheimer's Association, 2020) - have a high risk to progress to dementia. A recent systematic review reported a median conversation rate of 31.5% from MCI to Alzheimer's dementia over 5 years (Ward, Tardiff, Dye, & Arrighi, 2013). While Alzheimer's dementia and more vascular dementia are the most common types of dementia, many different types of dementia manifest with cognitive impairments: dementia with Lewy bodies, frontotemporal lobar degeneration, and dementias associated with traumatic brain injury, infections, Parkinson's disease, and alcohol abuse (Livingston et al., 2017). There are additional risk factors influencing cognitive abilities in normal aging, i.e. in the absence of primary neurodegenerative processes, such as human immunodeficiency virus (HIV), depression, and cardiovascular factors including diabetes, hypertension, obesity, and smoking (Livingston et al., 2017).

Taken together, dementias of different aetiologies represent one of the most urgent health challenges facing the world today, with an increasing associated social and economic burden. Decisively, dementia is not solely a memory problem, but can, also in early stages, be characterized by debilitating deficits in executive function, processing speed, comprehension, language, and mood (Kirova, Bayes, & Lagawart, 2015). Globally, the World Health Organization estimates 47 million people currently have dementia and this number is expected to triple by 2050 (WHO, 2017). This growing problem is in part due to accelerated ageing of the population, as well as the increasing prevalence of risk factors for dementia (Kuźma et al., 2018). Therefore, it is critically important to develop standardized neuropsychological assessment tools that are able to detect subtle cognitive impairments early on and with a high sensitivity. This will facilitate early diagnosis, clarification of comorbidities, identification of risk factors, potential treatments, and guidance of patients, as well as facilitate monitoring of cognition in healthy aging.

Detecting subtle cognitive impairments can be particularly challenging given that milder deficits tend to fluctuate in severity over time by impacting one specific cognitive domain whilst leaving other domains intact (Brambati et al., 2009). Our previous research has demonstrated that cognitive impairments acutely post stroke are commonly left undetected by brief dementia screening tools (Demeyere et al., 2016; Mancuso et al., 2018). Traditionally used global screening tools for dementia, such as the Mini-Mental State Examination (MMSE) (Folstein & McHugh 1975) and the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2015), take a binary view of cognition, relying on a summary score from subtasks with a single cut-off value for impairment, irrespective of age. Sometimes a broad-brush correction for education level is made by slightly adjusting the cut off. Consequently, the screens that take a binary approach are unable to detect subtle or domain-specific impairments due to the lack of subtest normative data and, frequently, the lack of population specific normative cut-offs (Robotham, Riss, & Demeyere, 2019; Wong et al., 2015). The MMSE and MOCA also contain many subtasks which are meant to assess non-language cognitive functions, but are heavily dependent on intact language function. For example, the MOCA's attention subtest requires participants to verbally repeat sequences of numbers (Nasreddine et al., 2015). This means that patients with a language deficit would appear to be impaired on this task, regardless of their underlying attentional capacity. This inability to separate distinct cognitive impairments is particularly problematic for stroke patients, even long term post stroke, where a large proportion of left hemisphere stroke patients will have continuing language impairments (Pedersen et al 1995). Similarly, the language component makes the screen less appropriate in populations with low literacy (Borson, Scanlan, Watanabe, Tu, & Lessig, 2005; Demeyere et al 2016; Mungas, Marshall, Weldon, Haan, & Reed, 1996; Mungas, Reed, Marshall, & González, 2000). Therefore, the dominance of language causes interpretation problems and leads to suboptimal tests.

New cognitive screening approaches have attempted to resolve this issue. Specifically, the Oxford Cognitive Screen (OCS) (Demeyere, Riddoch, Slavkova, Bickerton, & Humphreys, 2015) is a validated and normed standardized test that provides a cognitive profile for each stroke survivor covering five core cognitive domains (attention, language, memory, number and praxis). The OCS was designed to be as inclusive as possible, with many aphasia- and neglect-friendly subtests. It also adheres to the current National Institute for Health and Care Excellence (NICE) guidelines and is recommended for use in the Royal College of Physicians National Clinical Guideline for Stroke (Rudd, Bowen, Young, & James, 2017). Recently, the OCS has been shown to be more sensitive to cognitive deficits in acute post-stroke cohorts than both the MoCA and MMSE (Demeyere et al., 2016; Mancuso et al., 2018). While this approach is ideal for acute stroke settings, more detailed assessment is required to detect more subtle cognitive impairments long term and determine the severity of the deficits identified. In addition, multiple comorbidities including vascular small vessel diseases are common, and a long term cognitive decline may be related to these small vessel factors as well as (Lawrence et al 2015), or even instead of the sustained focal infarct (Østergaard et al. 2016).

Here, we present the validation of the Oxford Cognitive Screen Plus (OCS-Plus), a tablet-based cognitive screening tool developed as an extension of the OCS. The OCS-Plus was designed to provide sensitive cognitive screening of subtle cognitive impairments. The OCS-Plus contains a range of subtests designed to assess finer-grained, domain general cognition across memory, executive function, and attentional tasks. Similar to the OCS, the OCS-Plus aims to avoid undue loading of language requirements as the design emphasizes visual-oriented assessments. A previous iteration of the OCS-Plus has been translated and validated in a population of older adults in a low literacy and socioeconomic setting (Humphreys et al., 2017). Results of that study indicated that most OCS-Plus showed high

task compliance and good validity, improving the measurement of cognition with minimal language content, thereby avoiding total floor and ceiling effects present in other short cognitive screens. Importantly, it also showed clear and consistent age and education effects in a healthy population, demonstrating sensitive measurements of cognitive ageing.

The widespread adoption of tablet computers has facilitated the implementation of cost-effective computerized cognitive screening tools (Pew Research Center, 2018; Anderson & Perrin, 2017, Koo & Vizer, 2019). Computerized cognitive assessments present several important advantages over pencil-and-paper-based assessments, including the standardization of test administration, recording of more detailed response metrics, and automated scoring (Bauer et al., 2012; Koski et al., 2011). The OCS-Plus is available as an Android App and aims to employ these improved data collection capabilities to more accurately provide detailed profiles of cognition in ageing adults.

The purpose of the current study is to report standardised normative data and investigate the validity, internal consistency, and reliability of the OCS-Plus within a large group of neurologically healthy older adults. This psychometric validation is necessary to determine whether the OCS-Plus represents a useful method for detecting and differentiating between a range of subtle cognitive impairments within clinical populations.

Methods

Participants

A cohort of 320 neurologically healthy participants completed the OCS-plus subtests (for test by test sample sizes, see the normative tables). Participants were recruited in three different sites: Oxfordshire, UK (n=161), Coventry, UK (n =73) and Munich, Germany (n = 86). All participants were recruited through convenience sampling, whilst striving to include representative characteristics regarding level of education and age. All participants provided

written informed consent under local ethics (Oxford University ethics reference 'MSD-IDREC-C1-2013-209'; Coventry University ethics reference 'P33179' approved by Coventry University Research Ethics Committee (internally funded by Coventry University Pump-Prime Research Grant Scheme); Ludwig Maximilian University of Munich psychology department ethics committee reference '10_2015_*authorname*_b'). The demographic information for the complete cohort of participants is presented in Table 1.

By combining the cohorts, we provide a large normative data sample representative of participants in a broad Western European cultural context. Prior to combining these international cohorts, their compatibility was evaluated by comparing performance on each of the OCS-Plus subtests between groups. Participants in the UK and Germany performed similarly on all OCS-Plus subtests, with the exception of the Figure Copy Test in which the UK cohort was found to perform significantly better (Mann-Whitney test, German[n = 86] mean = 54.16, UK[n =229] mean = 55.72, U = 6779.5, p <.001). Given the difference between these two groups, UK and German data was therefore not combined for this measure when establishing normative data. On the basis of statistical power and having a larger pool to norm from, data from the larger UK cohort was used to establish impairment thresholds for the Figure Copy Task. For all other tasks, normative data from both UK and German participants was used. For full details on the comparisons between the UK and German cohorts, see Supplementary File 1.

Table 1

Demographic characteristics of the sample

			UK		German	
Characteristic	*N*	Observed	*n*	Observed	*N*	Observed
Age *M* (*SD*, range)	320	*62.66* (*13.75*, 23 - 99)	234	*60.51* (*14.91*, 23-99)	86	*68.49* (*7.26*, 50-81)
Education-Standard : High	316	84 : 232	230	51 : 179	86	33:53:00
Hand- R : L : A	318	265:19:00	232	211:18:03	86	85:01:00
Sex- M : F	320	176 : 144	234	128 : 106	86	48:38:00

Note. Some values are missing due to attrition over time (2014-2019), with four participants no longer available to be contacted for correct information. Standard education was differentially characterized for the German and UK samples due to variations in education scoring, whereby the German cohort were marked on further education or not, and the UK were marked on number of years in formal education including higher education. For the UK cohort on the basis of school running from 5-18, we classed standard education as <=12 years and high education as >=13 years. For the German cohort having a university degree separated the standard or high education sample.

OCS-Plus

The OCS-Plus comprises ten short tests and can be completed in 20-30 minutes. The validation of the tool was completed using a stand-alone application on Windows Surface Pro tablets developed using Matlab (The MathWorks Inc, 2012) and Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). The OCS-Plus tool has now been developed on an Android platform with data either locally removed at each start of session, or uploaded to a cloud server, dependent on user settings. This Android version

creates an automated report comparing performance to the normative data presented here. For access to the tool, please contact the Oxford University Innovations Healthoutcomes team.

A brief description of each task, the cognitive functions they aim to assess, and the order of administration is provided in Table 2. After each task, the examiner documents the condition of testing to flag any potential confounds, such as task interruptions or participant fatigue. Similarly, when a subtask is skipped, the reason for why the task was not attempted is recorded. This extra information subsequently aids the interpretation of the performance and report.

The OCS-Plus uses accuracy-based measurements where possible. This approach differs from other conventional neuropsychological assessments which use response time to quantify performance. A time-based scoring method generally assumes that healthy controls perform at ceiling, and this assumption does not always hold true (Kessels, 2019). Additionally, relying on time-based performance metrics is problematic for clinical populations containing participants who may exhibit response slowness for physical reasons, (e.g. motor weakness or muscle conductance) which may confound assessment of underlying cognitive deficits (Low, Crewther, Ong, Perre, & Wijeratne, 2017). It has also been suggested that older populations prioritize slower, more controlled performance over speed-based response strategies (Forstmann et al., 2011). For these reasons, the OCS-Plus employs accuracy-based outcome measures rather than response time-based metrics wherever appropriate. One exception to this approach is the OCS-Plus' measurement of processing speed, which is inherently a time-based metric. However, this measure still takes accuracy into account and is derived by dividing time taken by task accuracy. This proportional scoring method means that patients with slow performance and low task accuracy will be flagged as exhibiting abnormally slow processing speeds.

The OCS-Plus subtests were designed to be sensitive to an extensive range of variance in performance in healthy participants, instead of aiming to achieve ceiling performance in all neurologically healthy participants. All tests were designed to have low educational and language demands by using demonstrations and practice trials, short-high frequency words, and multimodal presentations. In addition, the design of the OCS-Plus includes an integrated code for translation and adaptation to other languages. The OCS-Plus app automatically scores and reports the outcome of the patient or respondent's performance. At the end of the session, the app automatically produces an online in-app report per task with clear indications of whether the participant is impaired compared to normative data. To provide a visual summary of the task and domain impairments, the OCS-Plus app presents an impairment wheel where impaired cognitive functions are visually represented (see Figure 1). This visualisation is similar to the OCS visual snapshot result (Demeyere et al., 2015) and is based upon normative data. All impairment thresholds are derived from age-specific normative data. For high quality images of each task see Supplementary file 2.

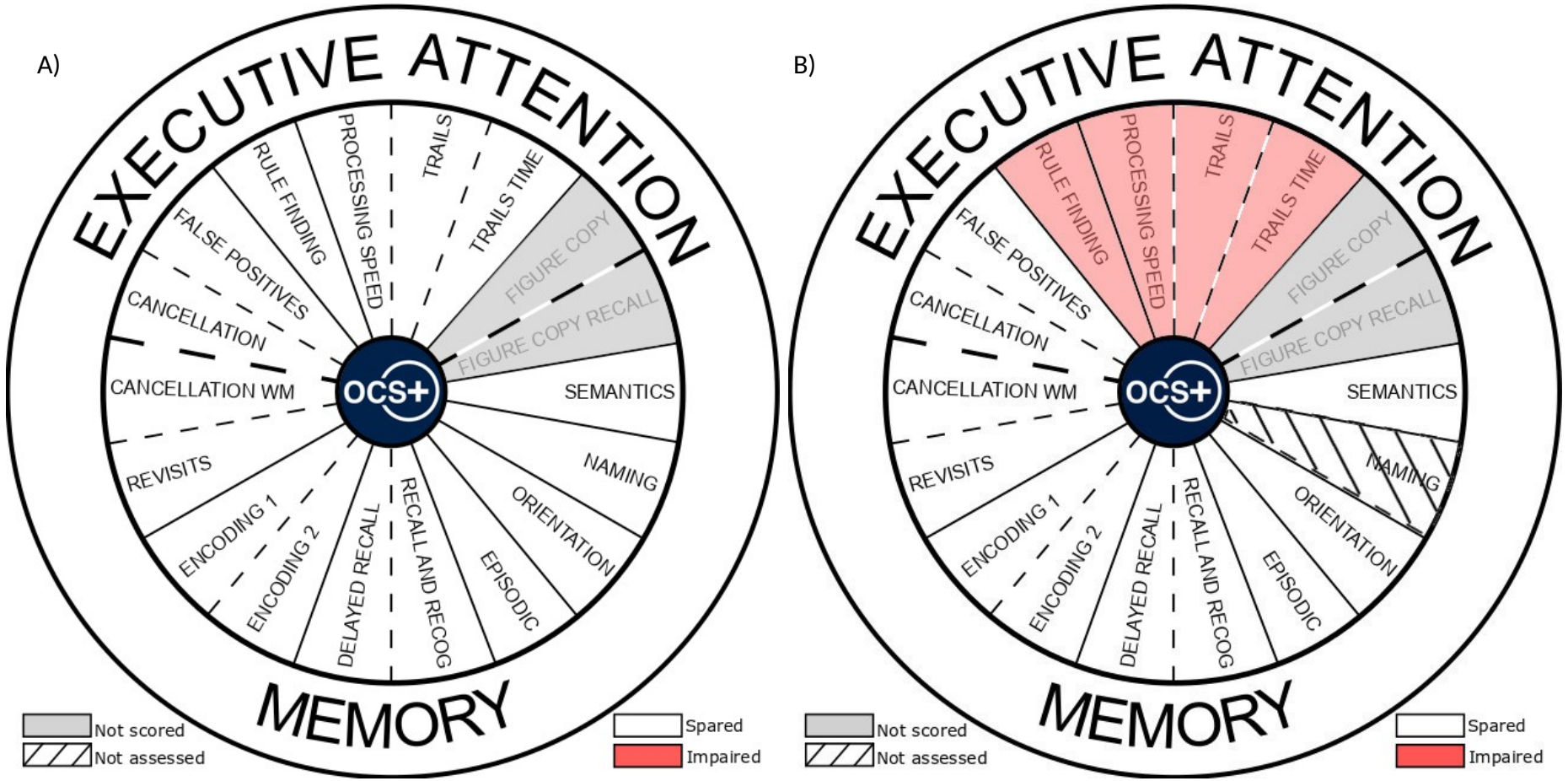

Figure 1. Part A) represents the base report outcome wheel and B) Represents a mockup of a patient who was impaired in the Trails and Rule Finding tasks compared to age matched normative data, they were not assessed on picture naming. Figure Copy and Figure Copy Recall are grey due to not having automatic scoring inbuilt in the app but instead requires transference of data to another app in order to score automatically.

Table 2.

Brief test descriptions and order of administration within OCS-Plus

Task Order	Task Name	Brief description	Scoring
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,	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.

Task Order	Task Name	Brief description	Scoring
		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.	
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. The effect of shape switching is assessed by taking performance on the switching task over performance on both baseline tasks summed.	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. 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	Five words are scored for correct free recall giving a score range from 0 – 5

Task Order	Task Name	Brief description	Scoring
		shown and read to the participant and they are asked to make a forced choice response.	
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	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

Task Order	Task Name	Brief description	Scoring
		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	error point per incorrect item selected in the visible cancellations task Correct revisits are scored as 1 error point per fruit item which is reselected in the invisible feedback conditions

Convergent and Divergent validity

The OCS-Plus was validated by comparing specific subtests to a series of analogous standardised neuropsychological tests in order to provide measures of convergent and divergent validity. See Table 3 for a visualisation of the specific comparisons which were conducted.

Addenbrooke's Cognitive Examination Revised (ACE-R). The ACE-R (Mioshi, Dawson, Mitchell, Arnold, & Hodges , 2006) is a short screening test designed to detect dementia-related cognitive impairment. The ACE-R was developed following the MMSE, which it incorporates, and requires approximately 15-20 minutes to complete. Performance on the ACE-R is quantified using a single total score out of 100 points (p) which is calculated by summing subtest scores across five domains: orientation and attention (18p), memory (26p), verbal fluency (14p), language (26p), and visuospatial processing (16p). The ACE-R impairment thresholds are determined using one of two standard total score cut-offs. Scores

of less than 82/100 represent significant impairment in cases where examiners wish to prioritise specificity over sensitivity, while scores of less than 88/100 are recommended when examiners aim to prioritize sensitivity over specificity.

Consortium to Establish a Registry for Alzheimer's Disease (CERAD). The CERAD-Plus test battery (Morris et al., 1989) measures cognitive performance in domains which are specifically impaired in Alzheimer's disease patients: memory, language, praxis, and orientation. This screening tool is able to differentiate between patients with mild and severe impairments and is therefore particularly useful for quantifying impairment severity and documenting the progression of cognitive decline over time. Furthermore, the CERAD-Plus has been found to have good objectivity, reliability, and validity measures and has been translated in numerous languages (Morris et al., 1989; Welsh, Butters, Hughes, Mohs, & Heymann, 1992). The CERAD-Plus contains semantic and phonemic verbal fluency tasks (Isaacs & Kennie, 1973; Spreen & Benton, 1977), the abbreviated Boston Naming Task (BNT; Kaplan, Goodglass & Weintraub, 1983), the MMSE (Folstein et al., 1975), the Word List Task (50p; Atkinson & Shiffrin, 1971; Rosen, Mohs & Davis, 1984; Mohs, Kim, Johns, Dunn & Davis, 1986), a visuospatial constructional praxis task, and the Trail Making Test (TMT; Reitan & Wolfson, 1995). These subtests are designed to assess a wide range of cognitive abilities including word retrieval, recognition, immediate and delayed recall, production, processing speed, cognitive flexibility, and executive function. However, this battery does not formally assess attention, though the TMT contains some attentional aspects (Reitan& Wolfson, 1995). Each CERAD-Plus subtest has been individually normed. This battery requires approximately 30-45 minutes to complete.

Rey–Osterrieth Complex Figure Test (ROCF). The ROCF is a visuospatial praxis test that draws upon various cognitive functions, including attention, visuospatial abilities, non-verbal memory, and task planning skills (Shin, Park, Park, Seol, & Kwon, 2006). This

task has three conditions: copy, immediate recall, and delayed recall. In the copy condition, subjects are presented with a complex line drawing and are asked to draw a copy of this figure from sight. In the immediate recall condition, the reference figure is removed and subjects are immediately instructed to draw the figure from memory. Finally, in the delayed recall condition, participants are asked to reproduce the figure from memory after a 30-minute delay period. Performance on the ROCF is scored according to the quantitative scoring system of Meyer and Meyer (1995), which includes 18 distinct figure elements which are separately scored with 0 to 2 points depending on correctness of position and completeness. Each figure reproduction is given a total score out of 36 possible points. This investigation only employs the copy and immediate recall conditions, as these conditions are most comparable with the OCS-Plus Praxis subtest. Participants are assigned a ROCF proportional score denoting the memory score as a percentage of the copy condition score for comparison with the OCS-Plus Figure Copy Recall score.

The Star Cancellation Test. The Star Cancellation Test is a visuospatial scanning task and part of the Behavioural Inattention Test (Wilson, Cockburn, & Halligan, 1987), a screening battery designed to assess the extent of hemispatial neglect. This task consists of a pseudorandom random search array of small and large stars, letters, and short words presented across a landscape A4 sheet of paper. Participants are instructed to search through this matrix and identify all small stars while ignoring all distractor stimuli. Participants are allowed 5 minutes to complete this task. Each participant is given a total score out of 56, representing the number of targets successfully identified. Neglect impairment is identified by considering spatial responses asymmetry.

Beck's Depression Inventory (BDI). The BDI is a standardized, self-report questionnaire that aims to assess the presence and severity of depression symptoms (Beck, Ward, Mendelson, Mock, & Erbaughet al., 1961). In this questionnaire, participants are

presented with a series of 21 Likert scale statements. Overall performance is scored by summing participant's Likert scale responses into a total score out of 63, with higher total scores representing a higher level of depressive symptoms. This measure is used to establish divergent validity of the OCS-Plus subtests to rule out the relation to depressive symptoms.

Table 3

Overview of Measures for Convergent and Divergent Validation

OCS-Plus subtask	Min	Max	Convergence task	Min	Max	Divergence task	Min	Max
Picture Naming	0	4	ACE Naming	0	12	BDI	0	63
			CERAD Boston Naming Task	0	27	ACE Visuospatial	0	16
Orientation	0	4	ACE Orientation	0	26	BDI		
						ACE Visuospatial		
Semantics	0	5	ACE Language	0	26	BDI		
			ACE Fluency	0	5	ACE Visuospatial		
Encoding 1			CERAD 1st Immediate Recall	0	9	BDI		
						ACE Visuospatial		
Encoding 2	0	5	CERAD 2nd Immediate Recall	0	10	BDI		
						ACE Visuospatial		
Delayed Recall	0	5	CERAD Delayed Recall	0	10	BDI		
						ACE Visuospatial		
Episodic Recognition	0	4	ACE Memory	0	26	BDI		
						ACE Visuospatial		
Processing speed	0	inf	CERAD Trail Making Test A	0	inf	BDI		
Executive score	0	100	CERAD Trail Making Test Time Ratio	0	inf	ACE Language	0	26
Cancellation	0	30	Star Cancellation Missing	0	56	BDI		
						ACE Language		
Invisible Cancellation	0	30	Star Cancellation Missing	0	56	BDI		
						ACE		

OCS-Plus subtask	Min	Max	Convergence task	Min	Max	Divergence task	Min	Max
Figure copy accuracy	0	60	ROCF Copy	0	36	Language BDI ACE		
Figure copy recall	0	60	ROCF 1st Recall	0	30	Language BDI ACE		
						Language		

Note. Min and Max values are not repeated for repeated validation tests.

Planned Analysis

First, we considered whether the two subsamples of healthy elderly participants from the UK and Germany were similar in their performance on the OCS-Plus before combining these groups into a single cohort.

The impairment threshold for each individual OCS-Plus subtest was then calculated based on the score distributions present within the healthy ageing control participant group. In this investigation, we employ impairment thresholds based on 5th centile (greater than two standard deviations*)* from the mean control performance following standard practice in neuropsychological testing (Loewenstein et al., 2006). Next, the reliability and validity of performance on the OCS-Plus subtests was evaluated. Task-specific correlations with established standardized measures were performed to provide evidence for convergent and divergent validity. Internal consistency was evaluated using Cronbach's Alpha as a measure of single factor internal reliability and test-retest validity was determined using Wilcoxon paired-samples tests.

Due to the division of cognitive domains in the OCS-Plus, an exploratory factor analysis using a Principle Axis method and Oblique rotation was conducted on subtest accuracy measures to assess the underlying components driving performance across various OCS-Plus subtests. This exploratory analysis enables greater confidence in the tests structure, which will need to be confirmed in patient populations. Kendall's tau correlation was

calculated to ensure that any variables were not highly inter-correlated. Next, parallel analysis was used to determine how many factors to retain. Next, oblique rotation was used to separate the factors but retain their potential underlying latent correlation. This methodology was chosen as the sub-scores of the OCS-Plus are not considered to be independent and some correlation is therefore expected. According to Guadagnoli and Velicer (1988), absolute sample size and the absolute magnitude of factor loadings determine reliable factor solutions. They argue that factors with a few low loadings should not be interpreted unless the sample size is 300 or more. As the sample size of the present study exceeds 300 participants, we will include factor loadings bigger than .3 and interpret all factors with more than one loading. Finally, we present one potential methodology for generating cumulative cognitive domain scores, which can be used to facilitate data interpretation within clinical settings. All analyses were performed in R version 3.5.1 (R Core Team, 2018) using packages such as *readxl* (Wickham & Bryan, 2019), *dplyr* (Wickham, François, & Henry, 2018) for data manipulation, *ggplot2* (Wickham, 2016) for plotting data, *Hmisc* (Harrell jr, 2018) for correlational data, *rcompanion* (Magiafico 2019) for computing Wilcoxon effect sizes, *sjstats* (Lüdecke, 2019) for eta effect size calculations, *psych* (Revelle, 2018) for the factor analysis, *GPArotation* (Bernaards & Jennrich, 2005) for the rotation of factors, and *reshape2* (Wickham, 2007) for other plots. The underlying raw data, the codebook containing distribution statistics on all variables, and the code used for all analyses reported are openly available through the Open Science Framework (anonymous view-only link https://osf.io/cfmwk/?view_only=7e5cdf8d963b46568a5069b50c3b4e76).

Results

Normative Data

Normative data and corresponding impairment cut-off thresholds for each of the OCS-Plus subtests were established. In line with previous cognitive assessment validation

techniques, different methods for determining impairment cut-offs were employed depending on the normalcy of subtest normative data (Demeyere et al., 2015). For subtests with non-normal distributions resulting from a restricted range of possible subtest scores, percentile-based impairment thresholds based on uncorrected sample score distributions were used. For all other subtests, cut-offs at 2 standard deviations below the control mean were employed. The normative data of OCS-Plus subtests and proposed cut-offs for impairment based on the full sample can be found in Table 4.

Table 4

Normative Data and Cut Offs for Impairment (Z-scores greater than 2SD from the mean or scores lower than 5th centile)

Task name	Measure	*N*	*M*	*SD*	median	min	max	centile
Picture Naming	Overall accuracy	320			4	1	4	3
Semantics	Overall accuracy	320			4	2	4	3
Orientation	Overall accuracy	320			4	2	4	3
Word recall	Encoding 1	320			5	1	5	3
	Encoding 2	320			5	0	5	4
	Delayed Recall accuracy	320			4	0	5	2
	Delayed Recall and recognition	320			5	3	5	4
Episodic Recognition	Episodic recall accuracy	319			4	1	4	3
Trails	Processing speed	320	33.83	18.70		11.02	143.38	
	Executive Score	320	81.68	25.04		0	100	
Rule Finding	Overall accuracy	320	26.74	8.20		3	43	
	Number of rules learned	320	2.99	1.32		0	5	
Figure copy	Overall accuracy	229	55.72	6.67		19	60	
	Figure recall accuracy							

Note. Means and *SD*s reported only for tasks with sufficient range in values. We proposed to use Z-score based impairments greater than 2 SDs from the mean. Measures with small ranges of possible scores are reported as median, min, max, and centiles. 5th centiles are chosen from determining cut offs for impairment. Asterisks reflect UK only norm data.

Trends of Performance across Age and Education

Cognitive abilities are not uniform across all age and education groups (Glisky, 2007). For this reason, the normative data sample was next split into subgroups and education- and age-specific impairment thresholds were established. Based on standard neuropsychological approaches, the sample was split into three age groups: below 60, 60-70, and above 70 years of age (Demeyere et al., 2015).

Participants were also divided into standard and high education groups. Within the UK cohort, participants with 12 or fewer years of education were considered to be within the standard education group and participants with more than 12 years were allocated to the high education group. These cutoffs are based on the UK legal education requirements, which stipulate that individuals must remain in school for at least 12 years. Due to different regulations between countries, German participants were binarised as standard education if they did not complete higher education, and classed as highly educated if they completed further education. See supplementary file 3 for tables of comparisons for age groups and education groups, as well as education and age specific cut offs.

Overall, participants with higher education levels and younger age tended to perform better on the OCS-Plus subtasks (Figure 2), yielding stricter impairment cutoffs for these groups than for older and standard education participants. Significant differences in performance were identified between various age and education groups, highlighting the need for age- and education-specific impairment thresholds on the OCS-Plus subtests. For visualization of cut offs per task for the age and education groups see Figure 2. Tables 5, 6, and 7 present normative data per education and age category across all tasks. There were suboptimal amounts of data to create generaliseable norms based off of education and age categories combined.

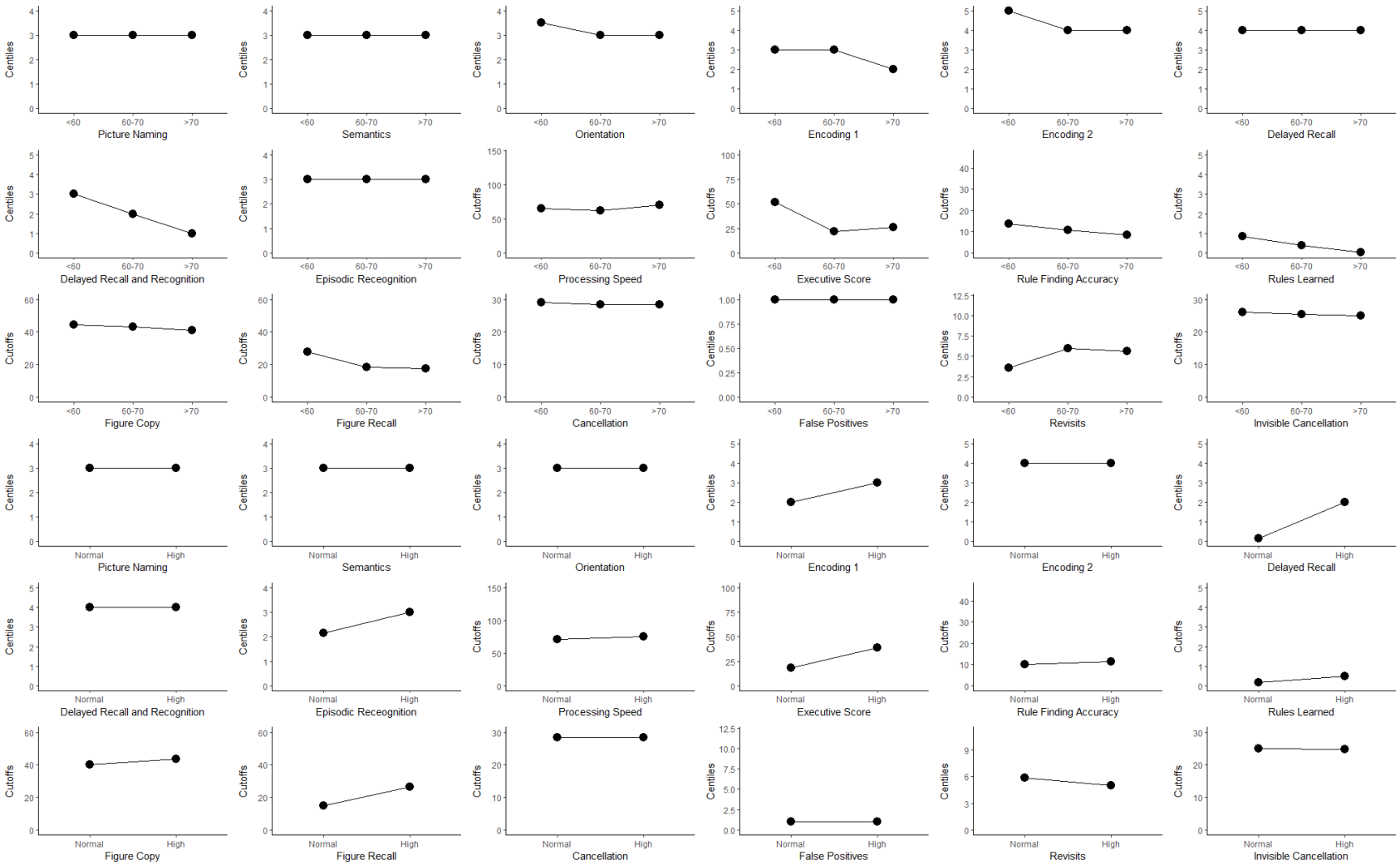
Centiles
Cutoffs
<60
60-70
>70
Normal
High
Picture Naming
Semantics
Orientation
Encoding 1
Encoding 2
Delayed Recall
Delayed Recall and Recognition
Episodic Receognition
Processing Speed
Executive Score
Rule Finding Accuracy
Rules Learned
Figure Copy
Figure Recall
Cancellation
False Positives
Revisits
Invisible Cancellation

Figure 2. Illustrates the trend for each of the categories (age or education) per subtask on the OCS-plus for the cut offs, whether centile or 2*SD* based, to increase or decrease dependent on higher age or lower education specifically on some task

Table 5
Normative Data and Cut Offs for Impairment per Education category (Z-scores Greater Than 2SD From the Mean or Scores Lower than 5th centile)

		Standard education							High education						
Task	Measure	*n*	*M*	*SD*	Med	Min	Max	Centile	*n*	*M*	*SD*	Med	Min	Max	Centile
Picture Naming	Accuracy	84			4	3	4	3	232			4	1	4	3
Semantics	Accuracy	84			4	3	4	3	232			4	2	4	3
Orientation	Accuracy	84			4	2	4	3	232			4	3	4	3
Word recall	Encoding 1	84			4	1	5	2	232			5	2	5	3
	Encoding 2 accuracy	84			5	0	5	4	232			5	0	5	4
	Delayed Recall accuracy	84			4	0	5	1	232			5	0	5	2
	Delayed Recall and recognition	84			5	4	5	4	232			5	3	5	4
Episodic Recognition	Episodic Recognition accuracy	84			3	2	4	3	231			4	1	4	3
Trails	Processing speed	84	34.27	18.71		15.81	98.34		179*	35.35	19.91		14	143	
	Executive Score	8	76.30	29.17		0	100		232	83.78	23.11		0	100	

Rule Finding	Accuracy	484	23.96	7.03	5	40	232	27.88	8.33	3	43
	Number of rules learned	84	2.57	1.20	0	5	232	3.15	1.33	0	5
Figure copy	Accuracy	82	53.88	6.83	33	60	176*	56.31	6.27	19	60
	Figure recall accuracy	83	39.04	12.05	13	59	228	45.29	9.52	15	60
Cancellation	Accuracy	84	29.66	.67	27	30	231	29.68	.62	27	30
	False positives	84	.08	.27	0	1	232	.07	.25	0	1
Invisible cancellation	Accuracy	84	28.32	1.69	23	30	231	28.49	1.87	13	30
	Correct revisits	84	1.63	2.09	0	12	231	1.33	1.82	0	11

Note: Means and *SD*s reported only for tasks with sufficient range in values. We proposed to use Z-score based impairments greater than 2 SDs from the mean. Measures with small ranges of possible scores are reported as median, min, max, and centiles. 5th centiles are chosen from determining cut offs for impairment. Asterisks reflect UK only norm data.

Table 6

Normative Data and Cut Offs for Centile Based Impairment per Age category (Scores Lower Than 5th centile)

		<60					*>60 & <70*					*>70*				
Task	Measure	*n*	Med	Min	Max	Cent	*n*	Med	Min	Max	Cent	*n*	Med	Min	Max	Cent

Picture Naming	Accuracy	111	4	1	4	3	101	4	3	4	3	108	4	2	4	3
Semantics	Accuracy	111	4	2	4	3	101	4	2	4	3	66	4	2	4	3
Orientation	Accuracy	111	4	3	4	3	101	4	3	4	3	66	4	2	4	3
Word recall	Encoding 1 accuracy	111	5	2	5	3	101	5	1	5	3	66	4	2	5	2
	Encoding 2 accuracy	111	5	4	5	5	101	5	0	5	4	108	5	3	5	4
	Delayed Recall accuracy	111	5	4	5	4	101	5	4	5	4	108	5	3	5	4
	Delayed Recall and recognition	111	5	0	5	3	101	4	0	5	2	108	4	0	5	1
Episodic Recognition	Episodic accuracy	110	4	1	4	3	101	4	2	4	3	108	4	2	4	3

Note: Cent refers to 5th centile of normative data.

Table 7

Normative Data and Cut Offs for Impairment per Age category (Z-scores Greater Than 2SD From the Mean)

		<60					*>60 & <70*					*>70*				
Task	Measure	*n*	*M*	*SD*	Min	Max	*n*	*M*	*SD*	Min	Max	*n*	*M*	*SD*	Min	Max
Trails	Processing speed	98*	31.44	16.78	16.53	117.92	101	31.67	15.05	15.56	98.34	108	39.54	22.58	14.24	143.38
	Executive Score	98*	88.48	18.31	28.57	100	101	77.33	27.90	0	100	108	81.72	25.40	0	100
Rule Finding	Accuracy	111	29.17	7.82	3	42	101	26.56	7.99	5	43	108	24.41	8.15	3	41
	Number of rules	111	3.34	1.25	0	5	101	2.98	1.30	0	5	108	2.64	1.31	0	5

	learned														
Figure copy	Accuracy	110	56.85	6.24	19	60	101	55.37	6.09	28	60	104	53.57	7.35	28	60
	Figure recall accuracy	110	47.02	9.70	19	60	101	42.43	11.27	15	60	104	41.25	9.88	13	59
Cancellation	Accuracy	111	29.82	.43	28	30	99	29.63	.62	28	30	107	29.56	.78	27	30
	False positives	111	.08	.27	0	1	100	.07	.26	0	1	108	.07	.26	0	1
Invisible cancellation	Accuracy	110	28.93	1.43	23	30	100	28.46	1.50	24	30	108	27.94	2.26	13	30
	Correct revisits	110	1.20	1.65	0	11	100	1.48	2.27	0	12	108	1.61	1.77	0	6

Note. The measures of correct revisits, false positives, and processing speed are impaired if an individual Z-score is greater than 2 *SD*s from the mean, whereas the remaining tasks are impaired if an individual's Z-score is lower than -2 *SD*s from the mean. Asterisks reflect UK only norm data.

Reliability

The reliability of each individual OCS-Plus subtest was evaluated using a split-half internal consistency measure and test-retest reliability analyses. The split-half approach was chosen as not all items or stimuli within a subtask are identical and, thus, random samples were taken to calculate the average internal consistency across many iterations, to partial out the effect of differences. We used 5,000 bootstrapped iterations to increase robustness of the result. The results of the analyses are presented in Table 8.

Table 8

Internal Split-half Consistency per task as Measured through Cronbach's Alpha

Measure	Alpha
Picture Naming	.23
Semantics	.17
Orientation	.04
Encoding 1	.41*
Encoding 2	.67*
Delayed Recall	.10
Delayed Recall and Recognition	1*
Incidental Episodic Memory	.06
Trails Baseline	.71*
Trails Switching	.87*
Rule Finding	1*
Cancellation	.96*
Invisible Cancellation	.96*
Figure Copy	.94*
Figure Recall	.93*

Note. * refer to reliability ratings we interpret due to the test assumptions being met. Trails baseline and switching conditions are included here to elude to reliability of the Executive score which has no trial level data due to being a ratio of baseline and switching performance.

Internal consistency per task was generally good with most Alpha values exceeding the standard threshold for good internal consistency (.70). However, a subset of OCS-Plus tasks was found to have lower Alpha values. Specifically, tasks with an inherently low total score variance resulting from a limited number of possible outcome scores (Picture Naming, Orientation, Semantics, Delayed Recall, and Recognition) were associated with low alpha values. Additionally, Incidental Episodic memory was found to have a low alpha value due to negative item correlations within the total score which negates higher alpha levels. Note that the negative item correlations were not present for Encoding 1 or Encoding 2. We report the alpha values for each measure for transparency, but, due to test assumptions, we interpret the values which could be computed without error (these are identified in the table with an asterisk).

Test-retest reliability. A group of 30 healthy ageing controls were retested on the OCS-Plus, on average 320 days apart (*SD*=265.89, range=30 – 1182), ensuring that they remained neurologically healthy at the second administration. Figure 4 presents plots of the test-retest reliability of this sample on the various OCS-Plus subtests. We found statistically stable test performances across time. Performance in a number of OCS-Plus subtests was near ceiling in the test-retest cohort. The resultant lack of variance precluded the calculation of correlation or intra-class correlation consistency for these subtests.

Consistency at the group level was then assessed comparing test-retest performance using a paired sample Wilcoxon test. This approach was chosen over equivalence testing as selecting a minimum effect size of interest would have been biased by the researcher's knowledge of the results of the previous analysis. The subtest test-retest analyses revealed that performance was not statistically different for any of the OCS-Plus tests before and after correction for multiple comparisons ($\boldsymbol{\alpha}_{\text{corrected}} = 0.003$, Table 9). Therefore, performance on the OCS-Plus subtests was not statistically different over time as a cohort.

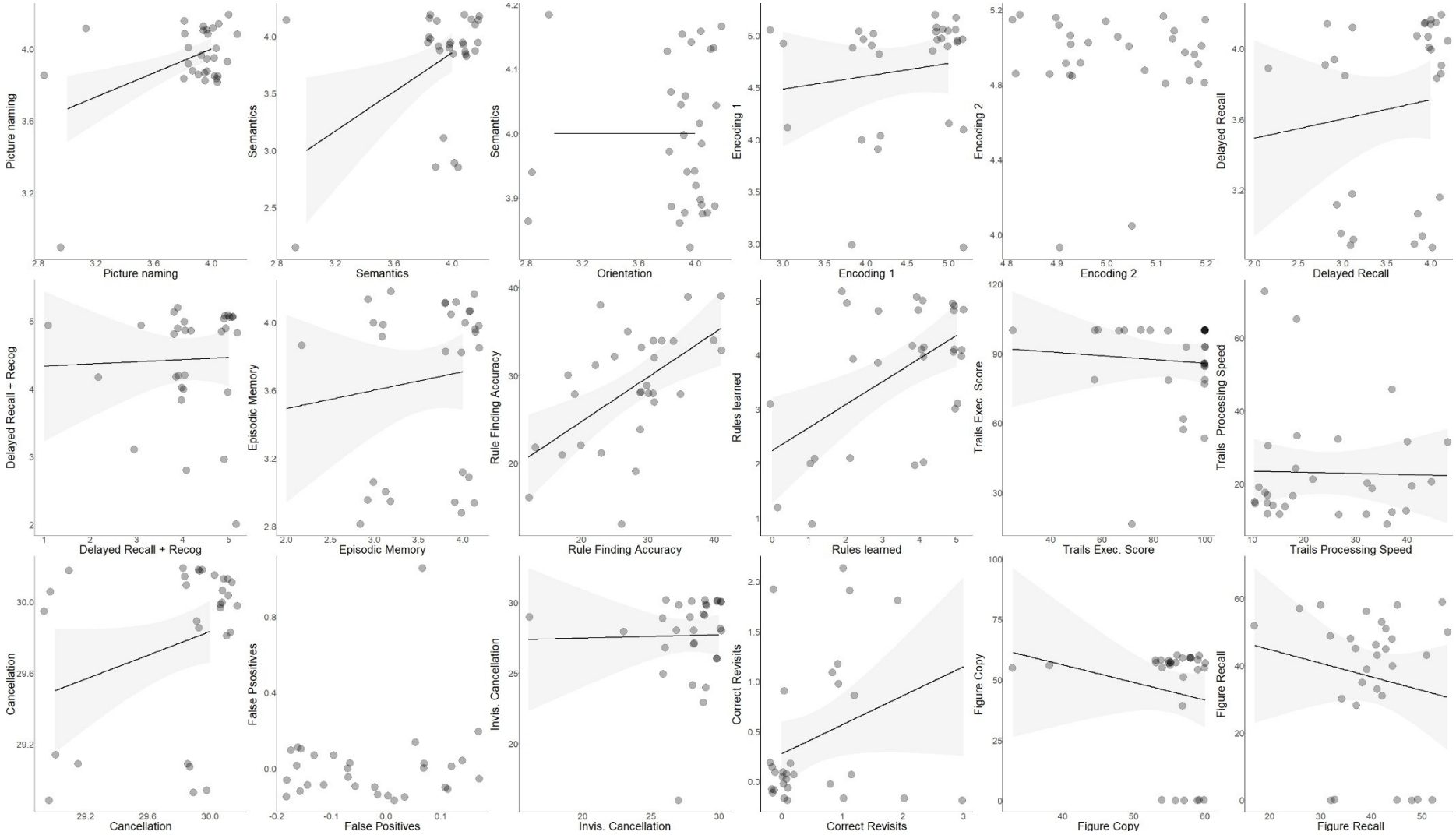

Figure 3. Illustrates the test retest reliability of the sample across time per task in the OCS-Plus (original administration is on the X axis and retested administration is on the Y axis), *R* values are Spearman's rho correlations. Jitter (spacing the data points) is imposed to aid visualization

of the overlapping data, therefore some values on the scales reflect jitter not real data values.

Table 9

Wilcoxon Paired-samples Test of Test-Retest Differences

Measure	Wilcox value	*p* value	*r*
Picture Naming	0	.35	0
Semantics	17.50	.13	-.07
Orientation	0	.15	0
Encoding 1	28.50	.22	.05
Encoding 2	3	.35	0
Delayed Recall	27.50	.63	0
Delayed Recall and Recognition	34	.24	-.05
Episodic Recognition	27.50	.63	0
Rule Finding Accuracy	182.50	.45	0
Rules Learned	82	.39	0
Trails Executive Score	139	.99	.02
Trails Processing Speed	266	.50	-.08
Cancellation Accuracy	16	.78	0
Cancellation False Positives	0	1	-
Invisible Cancellation Accuracy	149	.75	.03
Invisible Cancellation Revisits	27	.62	-.41
Figure Copy	264.50	.52	-.51
Figure Recall	230.50	.79	-.53

Note. We used the r statistic as an effect size for the Wilcox test. *r* could not be computed for false positives due to lack of variance.

Validity

A cohort of 86 German and 73 UK participants completed both the OCS-Plus and the battery of analogous standardized neuropsychological tests. The entire German cohort completed all of the ACE-R battery, 73 of the UK cohort completed the ACE III Picture naming. Data from this cohort was used to assess the validity of the OCS-Plus subtests. Convergent and divergent correlations between subtests of the OCS-Plus and the validation tests are summarized in Table 10. As a whole, the OCS-Plus subtests demonstrated good divergent validity versus theoretically unrelated constructs.

Table 10

Convergent and Divergent Validity Analysis of the OCS-Plus

	Convergent validity				Divergent validity			
OCS-Plus subtask	Task	*N*	*r*	*p*	Task	*N*	*r*	*p*
Picture Naming	ACE Naming	86	.14	.19	BDI	86	.05	.58
	CERAD Boston Naming Task	86	.10	.32	ACE Visuospatial	159	.12	.10
Orientation	ACE Orientation	86	.50	<.001*	BDI	86	.01	.94
					ACE Visuospatial	159	.00	.98
Semantics	ACE Language	159	.26	<.001*	BDI	86	.01	.96
					ACE Visuospatial	159	.13	.09
Encoding 1	CERAD 1st Immediate Recall	86	.16	.08	BDI	86	.01	.93
					ACE Visuospatial	159	.20	.005
Encoding 2	CERAD 2nd Immediate Recall	86	.24	.01	BDI	86	-.08	.41
					ACE Visuospatial	159	.17	.02
Delayed Recall	CERAD Delayed Recall	86	.35	<.001*	BDI	86	-.02	.79
					ACE Visuospatial	159	.10	.15
Episodic Recognition	ACE Memory	159	.06	.37	BDI	86	-.19	.04
					ACE Visuospatial	159	.07	33
Processing speed	CERAD Trail Making Test A	86	.21	.004	BDI	86	.11	.14
					ACE Memory	159	-.05	.42
Executive score	CERAD Trail Making Test Time Ratio	86	0	.98	BDI	86	.04	.57
					ACE Language	159	-.01	.84
Cancellation	Star Cancellation Missing	85	-.23	.03	BDI	85	.06	.54
					ACE Language	15	-.04	.63

Invisible Cancellation					BDI	9 86	.05	.58
					ACE Language	15 9	-.06	.33
Figure copy accuracy	ROCF Copy	86	.25	.005	BDI	86	-.02	.85
					ACE Language	15 7	.04	.51
Figure copy recall	ROCF 1st Recall	85	.17	.02	BDI	86	-.14	.08
					ACE Language	15 9	-.01	.82

Note. BDI is Becks Depression Index. ACE is Addenbrookes Cognitive Evaluation. CERAD is Consortium to Establish a Registry for Alzheimer's Disease, ROCF is Rey–Osterrieth Complex Figure Test. NA refers to no correlation possible due to lack of variance in ACE-Orientation. Processing speed is correlated with ACE-Memory as arguably ACE- Visuospatial and -Language are related given anecdotal reports of patients talking themselves through the rule to follow which may add time and visuospatial skills. To check our assumption we ran correlations between processing speed and both ACE-Visuospatial and –Language and both were small but significant. Asterisks represent alpha corrected significance (convergent = .05/13 comparisons, divergent = .05/26).

Family wise error rate corrections were used to correct for multiple comparisons when evaluating convergent validity. Given that 13 different convergence comparisons were made, a corrected alpha level of .003 was employed. The convergent validation analysis results revealed low, but statistically significant convergence correlations for most tasks pre-alpha correction. Only 3/13 OCS-Plus subtasks demonstrated significant associations with their primary validity measure after alpha correction. However, performance on several of the OCS-Plus subtests were not found to be significantly associated with analogous neuropsychological assessments (Table 10).

Exploratory Factor Analysis

An exploratory factor analysis was conducted on the OCS-Plus subtests to help elucidate the underlying mechanisms driving performance across the OCS-Plus as a whole. In this analysis, all 18 scored items were considered. First, Kendall's tau correlation was calculated to ensure that any variables were not highly inter-correlated and we observed a max correlation of $r = .84$ between rule finding accuracy and number of rules learned (see supplementary file 4). Parallel analysis (Horn, 1965) with oblique rotation yielded six significant underlying factors summarising the variance within OCS-Plus subtest performance. This methodology is particularly useful as it compares each potential factor eigenvalue to eigenvalues for the corresponding factor in many randomly generated data sets and only retains factors whose eigenvalues are greater than their randomly generated pairs (Watkins, 2006). In this analysis, Principle Axis algorithm was chosen as it is a conventional method of factor analysis (r function documentation found here https://www.rdocumentation.org/packages/psych/versions/1.8.12/topics/fa), and as we did not need to draw firm conclusions from it due to exploratory nature and possible different fractionation compared to clinical populations. We identified the factor number through

1,000 iterations of parallel analysis, as after 1,000 the factor number became stable. Before 1,000 there was a range of factor numbers below and above the structure identified due to very close eigenvalues below 1 which can occur in smaller samples.

Table 11 illustrates the factor loadings after oblique rotation. The items that cluster on the same components suggest an executive factor (factor 1), a visuospatial factor (factor 2), a delayed memory factor (factor 3), an attention factor dominated by the cancellation task (factor 4), a memory Encoding factor (factor 5), and finally a naming and semantic understanding factor (factor 6). In order of factor number the proportion of variance accounted for is as follows 1) 10.40%, 2) 7.70%, 3) 5.20%, 4) 4.20%, 5) 5%, and 6) 2.8%, cumulatively accounting for 35% of variance in the data.

Table 11

Exploratory Factor Analysis (Principle Axis Method) of the Structure of the OCS-Plus using Oblique Rotation

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Picture Naming	-	-	-	-	-	.46
Semantics	-	-	-	-	-	.44
Orientation	-	-	-	-	-	-
Encoding 1	-	-	-	-	.40	-
Encoding 2	-	-	-	-	.45	-
Episodic Recognition	-	-	-	-	-	-
Delayed Recall	-	-	.59	-	-	-
Delayed Recall and Recognition	-	-	.63	-	-	-
Rule finding	.90	-	-	-	-	-
Rules learned	.94	-	-	-	-	-
Trails Exec. Score	-	-	-	-	-	-
Processing Speed	-	-	-	-	-	-
Cancellation	-	-	-	.63	-	-
False positives	-.33	-	-	-	-	-
Invisible Cancellation	-	-	-	.48	-	-
Correct Revisits	-	-	-	-	-	-
Figure Copy	-	.64	-	-	-	-
Figure Recall	-	.89	-	-	-	-

Note. We only interpreted factors with more than one loading and with loadings greater than .3 and have excluded the pure time measure of executive time.

Other potential scoring methods

Lastly, one potential methodology for generating cognitive domain cumulative scores which can be used to facilitate data interpretation within clinical settings was explored. Based on the outcome of the factor analysis, six separate domain-specific scores were generated with scores weighted by factor loadings; executive function, praxis, delayed memory, attention, encoding, and naming and semantic understanding. Measures included in each score, normative data, and corresponding impairment thresholds for these domain total scores are presented in Table 12. This domain scoring system represents one of many potential more global scoring methods. Further research is needed to investigate the utility of the proposed alternative scoring methods.

Table 12

Normative Data of Scores and Cut Offs for Impairment (Scores lower than 5th centile and higher than 95th centile) based on factor structure

Summative Score	Included Measures	N	Med	Min	Max	Cent.	*M*	*SD*
Executive Function	Rule Finding Rules Learned False Positives	319	308.10	2.70	508.70	308	305.22	130.45
Praxis	Figure Copy Figure Recall	314	3749.38	1233.8	3893.4	3749	3581.95	428.01
Delayed Memory	Delayed Recall Delayed Recall and Recognition	320	5.51	1.89	6.10	6	5.52	.79
Attention	Cancellation Invisible Cancellation	316	32.67	23.25	33.3	33	32.35	1.07
Memory Encoding	Encoding 1 Encoding 2	320	4.25	1.60	4.25	4	3.94	.46
Picture Naming and Semantic Understanding	Picture naming Semantics	320	3.60	1.78	3.60	4	3.5	.24

Note. Each participant's raw score per measure is multiplied by the factor loading of that measure when summed to create the domain score.

Discussion

The purpose of this study was to present normative data for a novel, tablet-based cognitive assessment aiming to detect fine-grained impairments within ageing adults. Age- and education-specific cutoffs were established for each of the OCS-Plus subtests, based on data from a large cohort of neurologically healthy older adults. The validity of the OCS-Plus subtests was then evaluated against a series of analogous standard neuropsychological assessments. The OCS-Plus subtests were found to have good divergent validity. Similarly, performance on many OCS-Plus subtests was found to significantly correlate with performance on analogous standard measures. The OCS-Plus was found to have good test-retest reliability. Finally, factor analysis suggests that performance on this battery as a whole is driven by six independent factors, which may represent underlying cognitive domains.

As the population ages, the prevalence of subtle cognitive impairment and dementia-spectrum deficits is expected to increase dramatically (WHO, 2017). Detecting early signs of cognitive impairments may provide important prognostic information and aid in identifying optimal ways in assisting ageing adults remain independent in their activities of daily living (Bruscoli & Lovestone, 2004; Merriman et al., 2019). For this reason, it is crucial to develop new, more sensitive cognitive assessment tools to detect these subtle cognitive impairments. Given the recent rapid increase in affordable, user-friendly tablet computers, this new technological availability can be exploited to provide more detailed and sensitive standardized neuropsychological batteries in clinical and research environments. The OCS-Plus was developed to serve as a more sensitive screen for finer-grained, subtle cognitive impairments which are frequently missed by common, brief cognitive assessments (Robotham et al., 2019; Wong et al., 2015).

Normative Data

The diverse, international cohort of healthy ageing adults included in this investigation were collectively found to perform well on the OCS-Plus subtests. However, significant variance was present within the normative scores for most subtests, implying that not all healthy-ageing participants were found to perform at ceiling. For example, some OCS-Plus subtests (namely the Rule Finding Task) were found to be particularly challenging, even for control participants. The normative sample was found to have an average score of 26.74/46 on this task, with the best-performing participant achieving a score of 43. This variance greatly increases the utility of OCS-Plus subtest scores as it allows each new participant to be given a quantitative z-score denoting their deviation from the normative data mean as well as a qualitative impaired/unimpaired categorization. These z-scores are more sensitive to subtle changes over time, meaning that they can potentially be employed to document both normal and abnormal changes in cognition over time. Subtest z-scores are automatically calculated by the OCS-Plus program and are included in each patient summary report.

Control participants' performance, however, was found to reach ceiling on the OCS-Plus subtests with a comparatively small range of potential total score outcomes. The cohort of ageing adults was found to score the maximum number of points on the Picture Naming, Semantics, Orientation, and Episodic Recognition OCS-Subtests. This inherent lack of variance means that, by using data from these tasks alone, it remains difficult to differentiate slight, quantitative changes in ability, with the development of qualitative cognitive impairment. However, when these scores are considered in the broader context of OCS-Plus performance, they may still be useful for identifying change in cognitive abilities over time. Furthermore, these smaller ranges tasks could be used to identify where a patient has scored

low on the other tests but has otherwise normal intelligence and naming abilities, which could differentiate severe from slight impairments.

Performance on OCS-Plus subtests was found to be significantly different between various age and education groups. While this difference in performance between groups was not always large, it was present and statistically significant within the vast majority of OCS-Plus subtests. Therefore, age- and education-specific impairment thresholds have been proposed. These cut-offs can be applied to distinguish between normal and abnormal performance in a range of more specific subgroups. However, performance on the OCS-Plus subtests with a restricted range of outcome scores (e.g. Picture Naming) was not found to differ significantly between age and education groups. This finding agrees with the conceptualization of these specific subtests as qualitative rather than quantitative metrics, as neurologically healthy adults perform at ceiling on them.

Reliability

The OCS-Plus subtests were found to demonstrate good test-retest reliability at the group level, despite the wide range of test-retest intervals. *W* values for some subtests were low due to inherent low variance. However, performance on OCS-Plus subtests, overall, was found to be stable across time within this investigation's neurologically healthy ageing participant sample. Internal consistency per task was generally good with most Alpha values exceeding the standard threshold for good internal consistency (.70). However, reliability statistics were not able to be calculated for subtasks with restricted possible total scores as participant's scores were at ceiling. Collectively, the reliability analyses conducted in this study suggest that the OCS-Plus represents a reliable neuropsychological assessment battery.

Validity

The convergent and divergent validity of the OCS-Plus subtests was evaluated by comparing performance on these tasks to performance on analogous, standardized

neuropsychological measures. Type I error rate corrected significant convergent correlations were identified in 3/13 of the analyses run. However, the majority of these comparisons had comparatively low correlation coefficients, possible due to low variance in ceiling type performance of the control participants. This suggests that, while performance on the OCS-Plus subtests and analogous neuropsychological metrics is significantly associated, these tests are not exactly identical or had too few lower range scores to compute reliable estimates

(e.g., Picture Naming, Orientation, Semantics etc). For instance, differences in possible score ranges between the CERAD picture naming task and the OCS-Plus Picture Naming task, or word recall comparisons were exacerbated by ceiling performance on tasks. Some difference in performance between OCS-Plus and pen-and-paper is expected, as the stimuli, experimental design, and difficulty level are similar, but not identical across these assessments. For example between the CERAD trail making test and the OCS-Plus Trails, where the OCS-Plus version is a shape analogue whereas the CERAD is numeric-alphabetic in format.

As a whole, OCS-Plus subtests were found to have good divergent validity versus assessments aiming to test theoretically unrelated constructs. For each correlation plotted and to see the underlying distributions of data, see Supplementary file 5.

Exploratory Factor analysis

An exploratory factory analysis was performed to elucidate the underlying factor structure driving performance across the various OCS-Plus subtasks. This analysis suggests that six factors drive performance, and these factors seem to be analogous to established cognitive domains. Factor 1 was loaded by rule finding and rule learning performance, which are thought to represent measures of executive function (Van Den Berg et al., 2009). Factor 2 was loaded by performance on Figure Copy and Figure Recall, suggesting that this factor may represent visuospatial functions. Factor 3 included performance on the delayed memory

tasks, factor 4 related to performance on the attentional aspects of the Selection task, and factor 5 was loaded by performance on memory encoding tasks. Finally, factor six was loaded by performance on the Picture Naming and Semantic tasks, suggesting that this factor represents semantic knowledge and comprehension.

However, this analysis was intended to be purely exploratory and the authors do not suggest that this factor structure represents a ground truth measurement. It is likely that this factor structure may vary across different clinical cohorts. For example, in amnesic patients there might be a greater collapsed memory component, which also includes performance on tasks such as the selection task, as this subtest requires subjects to remember instructions as they complete it**.** Similarly, patients with visuospatial neglect might have an additional collapsed visuospatial attention factor which encompasses performance on all tasks involving spatially-presented stimuli (Selection, Trail Making, Figure Copy, and Rule Finding). Additional research, and subsequent confirmatory factor analysis is needed to confirm this potential underlying factor structure in a more diverse patient cohort.

Potential summative scores and clinical application

The OCS-Plus outputs a detailed, task-specific performance summary following the completion of each patient assessment. While this detailed data is particularly useful for research purposes, a more generalized scoring approach may be more practical within clinical settings. Existing neuropsychological assessment batteries such as the Oxford Cognitive Screen outputs task-specific scores, which can then be compiled to determine the presence or absence of impairments within a broader cognitive domain (Demeyere et al., 2015). For example, a patient exhibiting impairment within the OCS sentence recall, orientation, and episodic recall tasks could be more broadly classified as having an impairment within the domain of memory. These broader, cognitive domain impairment may provide a more practical method for clinicians attempting to interpret OCS-Plus data. We have, therefore,

suggested one potential method for combining test scores across cognitive domains and have provided normative data cut-offs for using this alternate scoring approach. This method is described as one of many potential alternative, clinician-focused OCS-Plus scoring systems. Future research is needed to investigate the utility of this proposed domain scoring system and to identify other informative alternate scoring methods.

Study limitations

It is well established that cognitive impairment is not a binary, categorical deficit, but rather represents a diverse spectrum of related problems. The OCS-Plus is not meant to provide a method for separating the spectrum of cognitive decline into arbitrary impairment classification groups. Instead, it is designed as a tool for collecting more detailed cognitive performance metrics for individual patients which can then be employed to inform clinical decision making. The boundaries distinguishing normal, age-related cognitive decline from abnormal cognitive deficits are not clearly established and the OCS-Plus is not intended as a tool for allocating patients to these clinical groups. The OCS-Plus is designed to help provide information about each patient's cognitive abilities, which can then be applied to inform clinical decision-making.

The OCS-Plus outputs a wide range of performance metrics, a subset of which were introduced and evaluated within this project. Most OCS-Plus subtests record detailed information including the x, y coordinates and timestamps of each participant response as well as audio recordings. These more complex performance metrics can be analyzed to provide a more detailed analysis of participant performance. For example, spatial search strategy could be quantified based on responses within the selection and figure copy task and this data could be analyzed to evaluate task planning and organizational abilities. Additional research is needed to explore these potentially informative extensions of OCS-Plus functionality.

The cohort used to establish OCS-Plus normative data included a diverse, international sample of healthy ageing adults. However, it would be advantageous to norm the OCS-Plus in a wider sample. Previous research has suggested that the OCS-Plus represents a valid measurement of cognitive abilities within low-literacy settings (Humphreys et al., 2017), but further normative data would be useful. As the OCS-Plus is intended for use in clinical populations, additional data summarizing the performance of clinical groups and the factor structure underlying this performance are needed.

Conclusion

This project presents normative data for the OCS-Plus and demonstrates the reliability and initial validity of this novel, tablet-based cognitive assessment battery in a neurologically healthy ageing cohort. This screening tool can be used to create informative summaries of finer-grained cognitive impairments in healthy ageing and clinical groups. Future research should aim to establish the feasibility and underlying factor structure of the OCS-Plus subtests in various samples of neurological patients, such as stroke, MCI, and dementia.

References

Anderson, M., & Perrin, A. (2017). Tech adoption climbs among older adults. Pew Research Center, 1-22.

Alzheimer's Association. (2020). *Mild Cognitive Impairment (MCI)*. Retrieved from https://www.alz.org/alzheimers-dementia/what-is-dementia/related_conditions/mild-cognitive-impairment

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225*(2), 82-91.

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the american academy of clinical neuropsychology and the national academy of neuropsychology. *The Clinical Neuropsychologist*, *26*(2), 177-196.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Arch Gen Psychiatry* 4; 561-571

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*(5), 676-696.

Borson, S., Scanlan, J. M., Watanabe, J., Tu, S., & Lessig, M. (2005). Simplifying detection of cognitive impairment: Comparison of the Mini-Cog and Mini-Mental state examination in a multiethnic sample. *Journal of the American Geriatrics Society*, *53*(5), 871-874.

Brainard, D. H. (1997). Spatial vision. The Psychophysics Toolbox, 10, 433-436.

Brambati, S. M., Belleville, S., Kergoat, M. J., Chayer, C., Gauthier, S., & Joubert, S. (2009). Single-and multiple-domain amnestic mild cognitive impairment: two sides of the same coin. *Dementia and geriatric cognitive disorders*, *28*(6), 541-549.

Bruscoli, M., & Lovestone, S. (2004). Is MCI really just early dementia? A systematic review of conversion studies. *International Psychogeriatrics*, *16*(2), 129-140.

Busse, A., Angermeyer, M. C., & Riedel-Heller, S. G. (2006). Progression of mild cognitive impairment to dementia: A challenge to current thinking. The British Journal of Psychiatry, 189(5), 399-404.

Demeyere, N., Riddoch, M. J., Slavkova, E. D., Bickerton, W., & Humphreys, G. W. (2015). The oxford cognitive screen (OCS): Validation of a stroke-specific short cognitive screening tool. *Psychological Assessment*, *27*(3), 883.

Demeyere, N., Riddoch, M. J., Slavkova, E. D., Jones, K., Reckless, I., Mathieson, P., & Humphreys, G. W. (2016). Domain-specific versus generalized cognitive screening in acute stroke. *Journal of Neurology*, *263*(2), 306-315.

Dichgans, M., & Leys, D. (2017). Vascular cognitive impairment. Circulation Research, 120(3), 573-591.

Finkel, D., Reynolds, C. A., McArdle, J. J., & Pedersen, N. L. (2007). Age changes in processing speed as a leading indicator of cognitive aging. *Psychology and Aging*, *22*(3), 558.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189-198.

Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: A structural model-based approach. *Journal of Neuroscience, 31*(47), 17242-17249.

Glisky, E. L. (2007). Changes in cognitive function in human aging. *Brain Aging: Models, Methods, and Mechanisms*, 3-20.

Guadagnoli, E., & Velicer, W. F. (1988). Stability of component patterns: A simulation study. *Psychological Bulletin*, *103*(2), 265-275.

Harrell, F. E. (2018). Hmisc: Harrell miscellaneous. R Package Version 4.2-0

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *32*, 179-185.

Humphreys, G. W., Duta, M. D., Montana, L., Demeyere, N., McCrory, C., Rohr, J., . . . Berkman, L., (2017). 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). *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *72*(1), 38-50.

Isaacs, B., & Kennie, A. T. (1973). The set test as an aid to the detection of dementia in old people. *The British Journal of Psychiatry*, *123*(575), 467-470.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). Boston Naming Test. Philadelphia, Lee and Febiger.

Kessels, R. P. (2019). Improving precision in neuropsychological assessment: Bridging the gap between classic paper-and-pencil tests and paradigms from cognitive neuroscience. *The Clinical Neuropsychologist*, *33*(2), 357-368.

Kirova, A. M., Bays, R. B., & Lagalwar, S. (2015). Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. *BioMed research international*, *2015*. https://doi.org/10.1155/2015/748212

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in psychtoolbox-3. Perception 36 ECVP Abstract Supplement.

Koo, B. M., & Vizer, L. M. (2019). Mobile technology for cognitive assessment of older adults: A scoping review. *Innovation in Aging*, *3*(1), igy038.

Koski, L., Brouillette, M., Lalonde, R., Hello, B., Wong, E., Tsuchida, A., & Fellows, L. K. (2011). Computerized testing augments pencil-and-paper tasks in measuring HIV-associated mild cognitive impairment. *HIV Medicine*, *12*(8), 472-480.

Kuźma, E., Lourida, I., Moore, S. F., Levine, D. A., Ukoumunne, O. C., & Llewellyn, D. J. (2018). Stroke and dementia risk: A systematic review and meta-analysis. *Alzheimer's & Dementia*, *14*(11), 1416-1426.

Lawrence, A. J., Brookes, R. L., Zeestraten, E. A., Barrick, T. R., Morris, R. G., & Markus, H. S. (2015). Pattern and rate of cognitive decline in cerebral small vessel disease: a prospective study. *PloS one*, *10*(8).

Livingston, G. (2017). Marriage and risk of dementia: systematic review and meta-analysis of observational studies.

Loewenstein, D. A., Acevedo, A., Ownby, R., Agron, J., Barker, W. W., Isaacson, R., ... & Duara, R. (2006). Using different memory cutoffs to assess mild cognitive impairment. *The American journal of geriatric psychiatry*, *14*(11), 911-919.

Low, E., Crewther, S. G., Ong, B., Perre, D., & Wijeratne, T. (2017). Compromised Motor Dexterity Confounds Processing Speed Task Outcomes in Stroke Patients. *Frontiers in neurology*, *8*, 484.

Lüdecke, D. (2019). sjstats: Statistical Functions for Regression Models. R PackageVersion 0.17.6.

Mangiafico, S. (2019). rcompanion: Functions to Support Extension Education Program Evaluation. R Package Version 2.3.7.

Mancuso, M., Varalta, V., Sardella, L., Capitani, D., Zoccolotti, P., Antonucci, G., & Italian OCS Group. (2016). Italian normative data for a stroke specific cognitive screening tool: The oxford cognitive screen (OCS). *Neurological Sciences*, *37*(10), 1713-1721.

Merriman, N. A., Sexton, E., McCabe, G., Walsh, M. E., Rohde, D., Gorman, A., ... & Horgan, F. (2019). Addressing cognitive impairment following stroke: systematic review and meta-analysis of non-randomised controlled studies of psychological interventions. *BMJ open*, *9*(2), e024429.

Meyers, J. E., & Meyers, K. R. (1995). Rey complex figure test and recognition trial professional manual Psychological Assessment Resources.

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The addenbrooke's cognitive examination revised (ACE-R): A brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry: A Journal of the Psychiatry of Late Life and Allied Sciences*, *21*(11), 1078-1085.

Mohs, R. C., Kim, Y., Johns, C. A., Dunn, D. D., & Davis, K. L. (1986). Assessing changes in Alzheimer's disease: Memory and language. In L. W. Poon, T. Crook, K. L. Davis, C. Eisdorfer, B. J. Gurland, A. W. Kaszniak, & L. W. Thompson (Eds.), *Handbook for clinical memory assessment of older adults* (p. 149–155). American Psychological Association. https://doi.org/10.1037/10057-012

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G. D. M. E., ... & Clark, C. (1989). The consortium to establish a registry for Alzheimer's disease (CERAD): I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*.

Mosti, C. B., Rog, L. A., & Fink, J. W. (2019). Differentiating Mild Cognitive Impairment and Cognitive Changes of Normal Aging. In *Handbook on the Neuropsychology of Aging and Dementia* (pp. 445-463). Springer, Cham.

Mungas, D., Marshall, S. C., Weldon, M., Haan, M., & Reed, B. R. (1996). Age and education correction of mini-mental state examination for english-and spanish-speaking elderly. *Neurology*, *46*(3), 700-706.

Mungas, D., Reed, B. R., Marshall, S. C., & González, H. M. (2000). Development of psychometrically matched english and spanish language neuropsychological tests for older persons. *Neuropsychology*, *14*(2), 209.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J.Am.Geriatr.Soc*. 4(53), 695-699.

Østergaard, L., Engedal, T. S., Moreton, F., Hansen, M. B., Wardlaw, J. M., Dalkara, T., ... & Muir, K. W. (2016). Cerebral small vessel disease: capillary pathways to stroke and cognitive decline. *Journal of Cerebral Blood Flow & Metabolism*, *36*(2), 302-325.

Pedersen PM, Jorgensen HS, Nakayama H, Raaschou HO, Olsen TS. Aphasia in acute stroke: incidence, determinants, and recovery. Ann Neurol. 1995; 38: 659–666

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437-442.

Pew Research Centre. (2019). Mobile fact sheet. Retrieved from http://www.pewinternet.org/fact-sheet/mobile/

R Core Team. (2018). R: A language and environment for statistical computing [computer software]. Vienna, Austria:

Rajan, K. B., Wilson, R. S., Weuve, J., Barnes, L. L., & Evans, D. A. (2015). Cognitive impairment 18 years before clinical diagnosis of Alzheimer disease dementia. *Neurology*, *85*(10), 898-904.

Reitan, R. M., & Wolfson, D. (1995). Category test and trail making test as measures of frontal lobe functions. *The Clinical Neuropsychologist*, *9*(1), 50-56.

Revelle, W. (2018). Psych: Procedures for Personality and Psychological Research. Version 1.8.12. Northwestern University, Evanston.

Robotham, R. J., Riis, J. O., & Demeyere, N. (2019). A danish version of the oxford cognitive screen: A stroke-specific screening test as an alternative to the MoCA. *Aging, Neuropsychology, and Cognition*, 1-14.

Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *The American Journal of Psychiatry, 141*(11), 1356–1364

Rudd, A. G., Bowen, A., Young, G. R., & James, M. A. (2017). The latest national clinical guideline for stroke. *Clinical Medicine*, *17*(2), 154-155.

Shin, M., Park, S., Park, S., Seol, S., & Kwon, J. S. (2006). Clinical and empirical applications of the Rey–Osterrieth complex figure test. *Nature Protocols*, *1*(2), 892.

Spreen, O., & Benton, A. L. (1977). Neurosensory center comprehensive examination for aphasia (NCCEA)—Revised. Victoria, British Columbia: University of Victoria Neuropsychology Laboratory.

The MathWorks inc. (2012). Matlab version 2012. Natick, MA.

Van Den Berg, E., Nys, G. M., Brands, A. M., Ruis, C., Van Zandvoort, M. J., & Kessels, R. P. (2009). The Brixton Spatial Anticipation Test as a test for executive function: validity in patient groups and norms for older adults. *Journal of the International Neuropsychological Society*, *15*(5), 695-703.

Ward, A., Tardiff, S., Dye, C., & Arrighi, H. M. (2013). Rate of conversion from prodromal Alzheimer's disease to Alzheimer's dementia: a systematic review of the literature. *Dementia and geriatric cognitive disorders extra*, *3*(1), 320-332.

Watkins, M. W. (2006). Determining parallel analysis criteria. *Journal of modern applied statistical methods*, *5*(2), 344-346.

Welsh, K. A., Butters, N., Hughes, J. P., Mohs, R. C., & Heyman, A. (1992). Detection and staging of dementia in alzheimer's disease: Use of the neuropsychological measures

developed for the consortium to establish a registry for alzheimer's disease. *Archives of Neurology*, *49*(5), 448-452.

WHO. (2019). Dementia. Retrieved from http://www.who.int/mediacentre/factsheets/fs362/en/

Wickham, H. (2007). Reshaping Data with the {reshape} Package. *Journal of Statistical Software*, *21*(12), 1-20.

Wickham, H. (2016). Ggplot2: Elegant graphics for data analysis Springer.

Wickham, H., & Bryan, J. (2019). readxl: Read Excel Files. R Package Version 1.3.1.

Wickham, H., Francois, R., & Henry, L. (2018). Dplyr: A Grammar of Data Manipulation.

Widera, E., Steenpass, V., Marson, D., & Sudore, R. (2011). Finances in the older patient with cognitive impairment: "He didn't want me to take over". *Jama*, *305*(7), 698-706.

Wong, A., Law, L. S., Liu, W., Wang, Z., Lo, E. S., Lau, A., . . . Mok, V. C. (2015). Montreal cognitive assessment: One cutoff never fits all. *Stroke*, *46*(12), 3547-3550.

Wilson, B., Cockburn, J., & Halligan, P. (1987). Development of a behavioral test of visuospatial neglect. *Archives of Physical Medicine and Rehabilitation*, *68*(2), 98-102.