Routledge
Taylor & Francis Group

Neuropsychological Rehabilitation
An International Journal

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pnrh20

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

Sam S. Webb & Nele Demeyere

Routledge
Taylor & Francis Group

CASE REPORT

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

Sam S. Webb[a] and Nele Demeyere [b]

[a]Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom; [b]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

ABSTRACT

We 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). Participants completed OxMET, MET-Home, Montreal Cognitive Assessment (MoCA), and questionnaires on activities of daily living, depression, technology usage, mobility, and disability. Forty-eight stroke survivors (mean age 69.61, 41.67% female, and average 16.5 months post-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 $r \geq .30$, $p < .006$). MET-Home accuracy was related to age ($B = -.04$, $p = .03$), 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 and interesting. 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.

ARTICLE HISTORY
Received 13 July 2023
Accepted 23 December 2023

KEYWORDS
Cognition; Stroke; Executive functions

Cognitive impairment after stroke impacts activities of daily living, from basic independence in self-care to instrumental activities such as work and participation in life (Maeir et al., 2011; Mole & Demeyere, 2020; Moore et al., 2021; Nys et al., 2006; Pohjasvaara et al., 2002). In particular, executive dysfunction has been found to be a core domain relating to functional outcomes (Laakso et al., 2019; Mole & Demeyere, 2020; Ownsworth & Shum, 2008). Executive

CONTACT Sam S. Webb sam.webb@psy.ox.ac.uk Department of Experimental Psychology, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock road, OX2 6GG

Supplemental data for this article can be accessed online at https://doi.org/10.1080/09602011.2024.2344326.

dysfunction is defined as a deficit in higher level cognitive abilities such as planning and self-monitoring behaviour, switching tasks or attention between tasks, and updating current schema to adjust behaviour (Goldstein et al., 2014; Miyake et al., 2000; Snyder et al., 2015). Accurate identification of executive dysfunction could reduce the burden on health services through early intervention (Josman et al., 2014). 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 (van der Zwaluw et al., 2011).

The Multiple Errands Test (MET) was originally developed as an ecologically valid and naturalistic assessment of executive functioning that involves completing errands within time and/or rule constraints (Shallice & Burgess, 1991). Clinicians have reported that they value the MET for its ecological validity and ability to reflect "real life" behaviour (Nalder et al., 2017). By including physical requirements, the MET also allows the assessment of a person's motor ability under a cognitive load – e.g., to assess their ability to walk while doing additional cognitive tasks. Furthermore, the MET offers clinicians flexibility, as specific task requirements can be adjusted dependent on rehabilitation goals (Nalder et al., 2017). Core metrics from the MET in general include the number of task completions and rules broken (Scarff et al., 2022). The MET has mixed psychometric evidence, with some research suggesting a strong association with the Dysexecutive Syndrome Questionnaire (DEX) as a measure of everyday dysfunction in executive function, but others reporting no such associations, though overall the MET is a strong contender as a measure of the impacts of executive dysfunction (Rotenberg et al., 2020).

However, labour and time costs to administer the MET are high, METs can take 15–40 min to administer (Nalder et al., 2017; Robnett et al., 2021); and they require clinician time to set up a local protocol version that is matched to MET literature and valid to compare to others (Scarff et al., 2022), to administer, score, and interpret the test. Importantly, the risk of injury in aging populations is high and expensive (Davis et al., 2010; Tian et al., 2013) and multiple errands tasks typically require walking through complex environments with a high cognitive load, which may lead to diverted attention and risk of falls (Koren et al., 2022; Nagamatsu et al., 2011). Though this falls risk is not unique to METs, it will be naturally be greater than for sedentary tasks.

Since its original inception, the MET has been revised and adapted for various contexts (hospitals, colleges, and shopping areas) and scenarios (e.g., skiing trips, errands around the home) (Alderman et al., 2003; Antoniak et al., 2019; Basagni et al., 2021; Burns et al., 2019; Cipresso et al., 2014; Clark et al., 2017; Cuberos-Urbano et al., 2013; Jovanovski et al., 2012; Knight et al., 2002, p. 2002; Maeir et al., 2011; Morrison et al., 2013; Rand et al., 2005; Rand & Katz, 2009; Raspelli et al., 2011; Robnett et al., 2021; Webb et al., 2021). Computerized adaptions of the MET have allowed for instant scoring, which reduces the time and safety cost analyses for administering a MET in clinical practice (Rand et al., 2009), as well as reducing examiner bias, ensuring more standardized and objective testing (Webb et al., 2021). These versions also significantly reduce the time and safety cost analyses for administering a MET in clinical practice (Rand et al., 2009). The psychometric properties of the MET are well described in a 2020 review, detailing that they largely meet convergence, divergence, sensitivity/specificity, and reliability benchmarks (Rotenberg et al., 2020).

The OxMET is a MET adaptation that has been developed for use on a computer tablet (Webb et al., 2021), 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 (Verstraeten et al., 2016). 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 5 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 (Webb et al., 2021). Convergence was above the convergence benchmark of $r = .30$ (Rotenberg et al., 2020). Furthermore, performance on the OxMET, in a subacute stroke rehabilitation context, reliably predicted short and longer term functional outcomes and activities of daily living over and above general cognitive scores from the Montreal Cognitive Assessment (Nasreddine et al., 2005; Webb & Demeyere, 2023). 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 research 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.). We 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.

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study (21 word solution by Simmons et al., 2012). Our manuscript adheres to the COSMIN guideline for studies on measurement properties (Gagnier et al., 2021) and the STROBE cohort checklist (von Elm et al., 2007).

Ethical approval procedures

Ethical approval was granted by the Medical Sciences Interdivisional Research Ethics Committee (REF: R73921/RE001) at the University of Oxford. All participants provided written informed consent.

Participant recruitment

The present study recruited chronic stroke survivors (approximately 6 months post stroke) from our 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 our 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 and Demeyere (2021) for a discussion on samples in neuropsychological research).

Sampling for both samples was based on convenience (those who replied to our 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 min (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 > 26 (Nasreddine et al., 2005) 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. We 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 (Shallice & Burgess, 1991) which requires participants to complete fourteen household tasks in their place of residence (Burns et al., 2019). 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 (Burns et al., 2019) and in Supplementary Table S1. In brief, the MET-home is scored based on the 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.

The OxMET

The OxMET is a simplified, standardized, normed, and validated computer-tablet version of the Multiple Errands Test (Shallice & Burgess, 1991) 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 (Webb et al., 2021). 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 = −10–10), as well as additional error metrics detailed in the original investigation (Webb et al., 2021) and in Supplementary Table S1.

Attempts and completion

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 (OxMET or MET-Home) 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.

Non-MET measures

We 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 we collected demographics and screened for global cognitive functioning with the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005). The MoCA is one of the most commonly used cognitive screening tools in stroke (Quinn et al., 2018) and scores range from 0–30, with higher scores reflecting better performance. The Geriatric Depression Scale (short form –15 item) (GDS-15) (Knight et al., 2004) was used to measure depression severity to account for the influence of mood on cognitive assessment and functional abilities (Green et al., 2021; Parikh et al., 1990), where higher scores reflect greater severity of depression. Basic Activities of Daily Life (ADL) performance was measured with the most commonly used (Quinn et al., 2011) functional independence measure, the Barthel Index (Mahoney & Barthel, 1965) (range of 0–20), and for instrumental ADLs we used the most commonly used measure in stroke trials (Salter et al., 2007): the Nottingham Extended Activities of Daily Living scale (NEADL) (Harwood & Ebrahim, 2002) (range = 0–66). For both ADL measures, higher scores reflect more independence. We additionally classified stroke severity at admission using the National Institute of Health

Stroke Scale (NIHSS) and disability/dependence using the Modified Rankin Scale (mRS) (Rankin, 1957) which is the most commonly used outcome measure in stroke trials (Broderick et al., 2017).

For the majority of participants ($n = 80$), we 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 (Roper, 2017). The scale on confidence in using technology is included in the technology screen that was adapted from an existing scale (Brumfitt & Sheeran, 1999). 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 our project we did not pre-select participants based on any criteria other than willingness to take part and eligibility. This means we selected stroke survivors who fit the inclusion criteria, and we 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 we only recruited those who could consent to take part.

Procedure

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

The duration of the research visit and full protocol was typically 1–1.5 hrs. Author SSW 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. Author SSW was trained and licensed to administer and interpret the MoCA. After completing each MET, we 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?" Author SSW recorded feedback on case record forms for each participant at the end of the session. Audio recording were not made. We 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, we present descriptive statistics and inferential statistics on sample differences in performance on each MET. Next, we correlated all OxMET to MET-Home metrics. Convergence overlap was determined as correlations above $r = .30$ where this benchmark is used in MET-literature (Rotenberg et al., 2020), and previously used in our OxMET studies (Webb et al., 2021; Webb & Demeyere, 2023). We examined the feasibility of administering both MET tasks: For all participants, we 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. We added time since stroke and stroke severity in separate identical models for stroke survivors. Missing data were not imputed and all analyses were conducted on complete data only. Missing data was sufficiently lower than that using only complete cases for each analysis that provided enough statistical power.

We briefly examined health economic considerations for using either task in clinical practice. We calculated staff costs per MET from the Schedule of Events Cost Attribution Template A calculation from the National Institute for Health and Care Research. We 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 we used pseudo-anonymized participant codes divided into either healthy controls or stroke survivors. We used a qualitative description approach to understanding and presenting the feedback data (Bradshaw et al., 2017). The approach aims to describe the experience under examination in a way that is easily understood and based primarily on direct quotes or statements (Bradshaw et al., 2017). We coded responses using a basic thematic analysis approach but, because we did not record responses verbatim in the first part of the study, we 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), we regressed accuracy to predict binarized positive or negatively valanced feedback.

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. We present the summary demographics of both healthy control and combined stroke survivor samples, as well as the percentage of missing data in Table 1. We ran Welch two sample *t*-tests to compare age and education demographics between samples for use as covariates in later analyses; we found a significant differences in education ($t(93.54)$, 2.43, $p = .01$, 95% CI = .32–3.16) but not age ($t(88.86)$, 0.72, $p = .472$, 95% CI = −3.25–6.96). We account for education differences in later group comparison analyses.

Descriptive statistics

In Table 2 we 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 (Burns et al., 2019; Webb et al., 2021). 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 which further justifies use as covariates in analyses.

Convergent associations between OxMET and MET-Home

We correlated all metrics from both METs using Spearman's Rho analysis due to the data type not being continuous and not normally distributed. The results of the convergent analyses are presented in Table 3 with correction for multiple comparisons ($** \, p < .00625$, $* \, p < .05$). All analyses are presented for total transparency, though we focus on like-for-like association (e.g., accuracy on OxMET and accuracy on MET-Home)

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-

Table 1. Summary of sample characteristics for all participants (healthy controls $n = 50$, stroke $n = 48$).

	Healthy controls		Stroke survivors	
Characteristic	*N*	Value	*N*	Value
Age (M (SD))	50 (100%)	71.46 (11.03)	48 (100%)	69.61 (14.14)
Education (M (SD))	50 (100%)	15.68 (3.32)	48 (100%)	13.94 (3.75)
Handedness	50 (100%)	*R* = 88%; *L* = 8%; *A* = 4%;	48 (100%)	*R* = 83.33%; *L* = 14.58%; *A* = 2.08%;
Sex	50 (100%)	*F* = 56%; *M* = 44%;	48 (100%)	*M* = 58.33%; *F* = 41.67%;
Ethnicity	50 (100%)	White: English, Welsh, Scottish, Northern Irish or British = 94%; White: Any other white background = 4%; Asian or Asian British: Indian = 2%;	48 (100%)	White: English, Welsh, Scottish, Northern Irish or British = 87.5%; White: Any other white background = 8.33%; 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%;
Days Since Stroke (M (SD, range))	–	–	48 (100%)	502.29 (547.47, 122–3,283)
Stroke type	–	–	48 (100%)	Ischaemic = 79.17%; Intracerebral haemorrhage = 12.50%; TIA = 6.25%; Subarachnoid haemorrhage = 2.08%;
Stroke side	–	–	48 (100%)	*R* = 47.92%; *L* = 33.33%; *B* = 12.50%; Undetermined = 6.25%;
Stroke Severity (Median (IQR))	–	–	42 (89%)	6 (4-9.5)
Mobility	50 (100%)	Walks completely independently with no difficulty = 96%; Walks independently but with difficulty (may use aid, e.g., frame/stick) = 4%	46 (98%)	Walks independently but with difficulty (may use aid, e.g., frame/stick) = 44.68%; Walks completely independently with no difficulty = 40.43%; Walks with assistance of 1 person = 10.64%; No sitting balance = 2.13%; Wheelchair + sitting balance = 2.13%
Modified Rankin Scale (M (SD))	50 (100%)	0 = 96%; 1 = 2%; 2 = 2%;	48 (100%)	0 = 8.33%; 1 = 18.75%; 2 = 27.08%; 3 = 33.33%; 4 = 10.42%; 5 = 2.08%;

Note: Percentage of complete data is presented in the "N" column in parentheses. *R* = right side, *L* = left side, *B* = bilateral, *A* = ambidextrous, *M* = male, and *F* = female.

Home, so they did not want to take part at all. Table 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.

Table 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	2.648	.107	.029
		Stroke	9.05	3.04					
	Frequency of Passes	Control	2.74	2.13	1	88	.296	.587	.003
		Stroke	2.90	2.08					
	Frequency of Rule Breaks	Control	8.10	4.32	1	88	.889	.348	.01
		Stroke	9.07	5.02					
	Inefficiencies	Control	.26	.60	1	88	.545	.462	.006
		Stroke	.32	.82					
	Number of Rules Broken	Control	2.30	.71	1	88	.34	.561	.004
		Stroke	2.39	.97					
	Omissions	Control	1.40	1.69	1	88	3.149	.079	.035
		Stroke	2.37	2.47					
	Partial Completions	Control	2.48	1.82	1	88	0.109	.743	.001
		Stroke	2.54	1.57					
	Rooms Visited	Control	7.08	3.44	1	82	2.682	.105	.032
		Stroke	5.68	3.01					
OxMET	Accuracy	Control	8.14	2.25	1	92	4.827	.031*	.05
		Stroke	6.42	3.66					
	Commissions	Control	.66	.96	1	92	0.464	.497	.005
		Stroke	.91	1.38					
	Frequency of Rule Breaks	Control	1.72	2.12	1	92	4.716	.032*	.049
		Stroke	3.33	3.61					
	Omissions	Control	1.02	1.45	1	92	7.332	.008*	.074
		Stroke	2.42	2.62					
	Partial Omissions	Control	.52	.79	1	92	1.965	.164	.021
		Stroke	.84	1.21					
	Perseveration	Control	.04	.20	1	92	1.063	.305	.011
		Stroke	.13	.50					
	Time	Control	174.09	53.87	1	92	5.635	.02*	.058
		Stroke	209.68	79.27					
	Total Error	Control	2.22	2.71	1	92	4.559	.035*	.047
		Stroke	4.18	4.49					
Non-MET	Barthel	Control	19.78	.62	1	95	9.701	.002**	.093
		Stroke	17.83	3.79					
	GDS15	Control	1.31	1.73	1	92	21.861	<.001**	.192
		Stroke	4.48	3.99					
	MoCA	Control	26.24	2.38	1	93	17.594	<.001**	.159
		Stroke	22.46	4.89					
	NEADL Total	Control	63.32	3.53	1	95	48.658	<.001**	.339
		Stroke	44.23	17.60					
	Tech Screen Total Used	Control	11.79	2.57	1	77	8.509	.005**	.1
		Stroke	9.32	3.81					
	mRS	Control	.06	.31	1	95	141.522	<.001**	.598
		Stroke	2.25	1.19					

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.

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

Table 3. Spearman's Rho correlations between OxMET and MET-Home performance metrics inclusive of healthy controls and stroke survivors.

	OxMET metrics							
MET-Home Metrics	Time	Accuracy	Omissions	Partial Omissions	Commissions	Perseveration	Frequency of Rule Breaks	Total Errors
Planning time	.41**	−.10	.12	−.02	.10	−.01	.11	.06
Time	.25*	.23*	−.25*	−.30**	−.20	.05	−.25*	−.28*
Accuracy	−.19	.37**	−.37**	−.26*	−.26*	−.01	−.36**	−.36**
Partial Completions	.07	−.07	.10	.06	.05	−.08	.08	.08
Omissions	.12	−.38**	.35**	.24*	.24*	−.02	.35**	.35**
Number of Rule Breaks	.03	.08	−.11	−.24*	−.13	−.11	−.13	−.15
Frequency of Rule Breaks	.18	−.07	.04	−.15	−.01	.06	.04	.00
Frequency of Passes	.24*	−.12	.11	.09	.10	−.03	.09	.09
Inefficiencies	.00	.02	.01	−.12	−.03	−.11	−.01	−.07
Rooms Visited	−.19	.25*	−.26*	−.29*	−.17	−.22*	−.26*	−.29*

Note: $** p < .00625$, $* p < .05$.

Table 4. Summary of attempt and completion of the OxMET and MET-Home.

	MET-Home		
OxMET	Not attempted	Attempted and not completed	Attempted and completed
Not attempted	1 (1.02%)	0 (0%)	1 (1.02%)
Attempted and not completed	0 (0%)	0 (0%)	0 (0%)
Attempted and completed	4 (4.08%)	3 (3.06%)	89 (90.82%)

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 7.5 inches.

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 down stairs 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.

Predicting performance completion and accuracy

Our 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, $B = -.04$, $p = .03$, sex, $B = -.98$ (males worse), $p = .03$, mRS, $B = -.63$, $p = .04$, 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, $B = -.20$, $p = .017$ 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 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 3.19 min versus 11.21 min 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 (Ambrens et al., 2022), 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

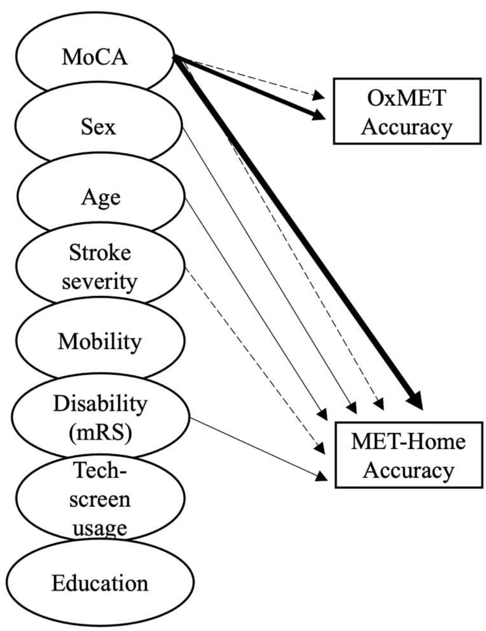

Figure 1. Visualization of statistically significant predictive associations across all participants (solid lines) and only stroke survivors (dashed lines) for task accuracy. Thicker lines are associations which survived corrections for multiple comparison ($p < .006$). Figure available under CC-BY 4.0 attribution license (https://osf.io/xw3s8).

participants who are at a falls risk, or who may not be, but do fall or otherwise injure themselves whilst maneuvering around the home. In this project we observed one participant rushing during the task and almost falling down the stairs, but who 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 (Tian et al., 2013). At least a third of older adults in the UK experience falls (Davis et al., 2010) and stroke survivors are at an even higher increased risk, with 40–70% of stroke survivors experiencing a fall each year (Kannan et al., 2019), and 14% experiencing one within the first month post-stroke (Wei et al., 2019).

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 Supplementary Table S3. Feedback was not recorded for eight participants due to examiner error. Following a summary of the feedback, we 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 valanced 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 were only fully visible once selected.

Association of MET accuracy to qualitative feedback given

We 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. We 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, we 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, $p = .26$).

Discussion

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

When accounting for sample differences in education, we only found differences in performance on the OxMET, not MET-Home. No significant differences were found below a Bonferroni corrected alpha level of .006 for MET-Home or OxMET. This contrasts with the original OxMET paper which found a larger effect size and more group differences previously, even when accounting for age and education (Webb et al., 2021). In Webb et al. (2021), they had over 200 participants and may have had more statistical power to detect effects than we did with 98 participants in total. One explanation is that the original OxMET study included controls which were much younger and more highly educated than the stroke cohort, and the stroke cohort where much more chronic in their stroke recovery. The current stroke and healthy sample were much better matched in age which may explain the differences in results here. It is not the case that not finding group differences that survive alpha correction makes the OxMET a bad test, but that there may be differences, but we had too small a sample to detect them (indicated by a significance below .05 but above our cut-off of .006, and moderate effect sizes).

Burns et al. (2019) did not account for age and education effects and found multiple group differences. We are unable to check 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 a well-matched study accounting for age and education may attenuate differences due to the known association of age and education with cognition (Demeyere et al., 2021).

We 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 our corrected alpha level too. Notably, most convergence was above the pre-defined alpha level of $r = .30$ (Demeyere et al., 2021; Rotenberg et al., 2020; Webb et al., 2021). The exception being MET-Home partial omissions and inefficiencies which did not relate 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. We 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), we are 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 (Burns et al., 2019).

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 (Clark et al., 2017), though similar to the present study, Clark et al. (2017) 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 (Juslin & Olsson, 2005; Schönbrodt & Perugini, 2013).

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. We did not find any significant independent predictors of completion of either the MET-Home or the OxMET, due to the 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 we 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 (Webb et al., 2021), though this was not in the context of other factors (e.g., it was a sole correlate).

In our brief examination of health economics we 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) (Verstraeten et al., 2016). 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 fundamental goal of OxMET is to assess executive functioning abilities which may affect everyday life, rather than to directly capture everyday life behaviours. As a very brief screening tool the OxMET has been shown to have predictive validity regarding future functional outcomes (Webb & Demeyere, 2023). 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.

We finally examined the feedback from participants on both tasks using a basic qualitative description approach and thematic analysis. Largely we found that there were more negative comments regarding the MET-Home, as opposed to the OxMET, especially surrounding the stress of the MET-Home. We interpret the findings of the qualitative examination of feedback to reflect the acceptability of the METs. We infer that there is less acceptability for the MET-Home as it can be stressful and not appropriate for those with low mobility. We take this interpretation from over 21 references to not wanting to complete item one of the MET-Home. Note, we discussed item one with the MET-Home authors, who asserted that this item cannot be removed or replaced with a non-call option. We 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. We 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 our high rates of OxMET attempts/completion.

Limitations of the study

While we were statistically powered to detect our benchmark correlation of $r \geq .30$, we added justifiable and important covariates of interest which may have reduced our statistical power to detect effects. We were one participant away from the minimum sample size required. Additionally, only 48 participants were stroke survivors, limiting our generalizability to stroke survivors. We 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. We collected feedback from participants after completing each MET, and the feedback was not recorded verbatim for all. We therefore describe the feedback only descriptively and did not conduct a full qualitative analysis. Initially, we recorded feedback that we 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. We realized the importance of collecting verbatim information during the project and adjusted our methods to do so for the remainder of the project. As such, some meaning may have been lost in our records, though we 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 (Rand & Katz, 2009) or Virtual Mall task (Rand et al., 2005). 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 Computerized Multiple Elements Test (Hynes et al., 2015) or Computerized Wisconsin Card Sorting Test (Tien et al., 1996) 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.

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.

Note

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.

Declarations

Nele Demeyere is a developer of the "OxMET" but does not receive any remuneration from its use.

Acknowledgements

We would like to thank all participants who gave us their time to aid in our research project. We would also like to thank MSc student Robyn Charleston, and research assistant Emma

Colbourne who helped with some of the data collection, and Georgina Hobden and Dr Suzanne Burns for helpful comments on an earlier draft of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by a fellowship award from the Stroke Association to S. S Webb (PGF 21100015). Nele Demeyere, (Advanced Fellowship NIHR302224) is funded by the National Institute for Health and Care Research (NIHR). The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR, NHS or the UK Department of Health and Social Care.

Data availability statement

The data that support the findings of this study are openly available in Open Science Framework at http://doi.org/ 10.17605/OSF.IO/PZWGF.

ORCID

Nele Demeyere http://orcid.org/0000-0003-0416-5147

References

Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society*, *9*(1), 31–44. https://doi.org/10.1017/S1355617703910046

Ambrens, M., van Schooten, K. S., Lung, T., Clemson, L., Close, J. C. T., Howard, K., Lord, S. R., Zijlstra, G. A. R., Tiedemann, A., Valenzuela, T., Vandelanotte, C., Chow, J., McInerney, G., Miles, L., Woodbury, A., & Delbaere, K. (2022). Economic evaluation of the e-health StandingTall balance exercise programme for fall prevention in people aged 70 years and over. *Age and Ageing*, *51*(6), afac130. https://doi.org/10.1093/ageing/afac130

Antoniak, K., Clores, J., Jensen, D., Nalder, E., Rotenberg, S., & Dawson, D. R. (2019). Developing and validating a big-store multiple errands test. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02575

Basagni, B., Bosetti, S., Cantelli, S., Ermon, S., Rovere, F., Navarrete, E., De Tanti, A., & Saviola, D. (2021). Development of a generic version of the multiple errands test for severe acquired brain injuries. *Applied Neuropsychology: Adult*, *0*(0), 1–8. https://doi.org/10.1080/23279095.2021.1990928

Bradshaw, C., Atkinson, S., & Doody, O. (2017). Employing a qualitative description approach in health care research. *Global Qualitative Nursing Research*, *4*, 2333393617742282. https://doi.org/10.1177/2333393617742282

Broderick, J. P., Adeoye, O., & Elm, J. (2017). Evolution of the modified Rankin scale and its use in future stroke trials. *Stroke*, *48*(7), 2007–2012. https://doi.org/10.1161/STROKEAHA.117.017866

Brumfitt, S. M., & Sheeran, P. (1999). The development and validation of the visual analogue self-esteem scale (VASES) 1. *British Journal of Clinical Psychology*, *38*(4), 387–400. https://doi.org/10.1348/014466599162980

Burns, S. P., Dawson, D. R., Perea, J. D., Vas, A., Pickens, N. D., & Neville, M. (2019). Development, reliability, and validity of the multiple errands test home version (MET–home) in adults with stroke. *American Journal of Occupational Therapy*, *73*(3), 7303205030p1–7303205030p10. https://doi.org/10.5014/ajot.2019.027755

Cipresso, P., Albani, G., Serino, S., Pedroli, E., Pallavicini, F., Mauro, A., & Riva, G. (2014). Virtual multiple errands test (VMET): A virtual reality-based tool to detect early executive functions deficit in Parkinson's disease. *Frontiers in Behavioral Neuroscience*, *8*, 405. https://doi.org/10.3389/fnbeh.2014.00405

Clark, A. J., Anderson, N. D., Nalder, E., Arshad, S., & Dawson, D. R. (2017). Reliability and construct validity of a revised Baycrest Multiple Errands Test. *Neuropsychological Rehabilitation*, *27*(5), 667–684. https://doi.org/10.1080/09602011.2015.1117981

Cuberos-Urbano, G., Caracuel, A., Vilar-López, R., Valls-Serrano, C., Bateman, A., & Verdejo-García, A. (2013). Ecological validity of the multiple errands test using predictive models of dysexecutive problems in everyday life. *Journal of Clinical and Experimental Neuropsychology*, *35*(3), 329–336. https://doi.org/10.1080/13803395.2013.776011

Davis, J. C., Robertson, M. C., Ashe, M. C., Liu-Ambrose, T., Khan, K. M., & Marra, C. A. (2010). International comparison of cost of falls in older adults living in the community: A systematic review. *Osteoporosis International*, *21*(8), 1295–1306. https://doi.org/10.1007/s00198-009-1162-0

Demeyere, N., Haupt, M., Webb, S. S., Strobel, L., Milosevich, E., Moore, M. J., Wright, H., Finke, K., & Duta, M. (2021). Introducing the tablet-based Oxford cognitive screen-plus (OCS-plus) as an assessment tool for subtle cognitive impairments. *Scientific Reports*, *11*(8000). https://doi.org/10.1038/s41598-021-87287-8

Gagnier, J. J., Lai, J., Mokkink, L. B., & Terwee, C. B. (2021). COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Quality of Life Research*, *30*(8), 2197–2218. https://doi.org/10.1007/s11136-021-02822-4

Goldstein, S., Naglieri, J. A., Princiotta, D., & Otero, T. M. (2014). Introduction: A history of executive functioning as a theoretical and clinical construct. In Goldstein S. & Naglieri J.A. (Eds.), *Handbook of executive functioning* (pp. 3–12). Springer.

Green, S. L., Gignac, G. E., Watson, P. A., Brosnan, N., Becerra, R., Pestell, C., & Weinborn, M. (2021). Apathy and depression as predictors of activities of daily living following stroke and traumatic brain injuries in adults: A meta-analysis. *Neuropsychology Review*, 1–19. https://doi.org/10.1007/s11065-021-09501-8

Harwood, R. H., & Ebrahim, S. (2002). The validity, reliability and responsiveness of the Nottingham extended activities of daily living scale in patients undergoing total hip replacement. *Disability and Rehabilitation*, *24*(7), 371–377. https://doi.org/10.1080/09638280110101541

Hynes, S. M., Fish, J., Evans, J. J., & Manly, T. (2015). Developing a computerised multiple elements test for organisational difficulties. *International Journal of Developmental Science*, *9*(2), 85–94. https://doi.org/10.3233/DEV-140157

Josman, N., Kizony, R., Hof, E., Goldenberg, K., Weiss, P. L., & Klinger, E. (2014). Using the virtual action planning-supermarket for evaluating executive functions in people with stroke. *Journal of Stroke and Cerebrovascular Diseases*, *23*(5), 879–887. https://doi.org/10.1016/j.jstrokecerebrovasdis.2013.07.013

Jovanovski, D., Zakzanis, K., Campbell, Z., Erb, S., & Nussbaum, D. (2012). Development of a novel, ecologically oriented virtual reality measure of executive function: The multitasking in the city test. *Applied Neuropsychology: Adult*, *19*(3), 171–182. https://doi.org/10.1080/09084282.2011.643955

Juslin, P., & Olsson, H. (2005). Capacity limitations and the detection of correlations: Comment on kareev (2000). *Psychological Review*, *112*(1), 256–267; discussion 280. https://doi.org/10.1037/0033-295X.112.1.256

Kannan, L., Vora, J., Bhatt, T., & Hughes, S. L. (2019). Cognitive-motor exergaming for reducing fall risk in people with chronic stroke: A randomized controlled trial. *NeuroRehabilitation*, *44*(4), 493–510. https://doi.org/10.3233/NRE-182683

Knight, C., Alderman, N., & Burgess, P. W. (2002). Development of a simplified version of the multiple errands test for use in hospital settings. *Neuropsychological Rehabilitation*, *12*(3), 231–255. https://doi.org/10.1080/09602010244000039

Knight, R. G., McMahon, J., Green, T. J., & Skeaff, C. M. (2004). Some normative and psychometric data for the geriatric depression scale and the cognitive failures questionnaire from a sample of healthy older persons. *New Zealand Journal of Psychology*, *33*(3), 163–170.

Koren, Y., Rozenfeld, E., Elefant, I., Khir, N., Glassberg, E., & Batcir, S. (2022). Does cognitive loading interfere with walking control? *Gait & Posture*, *96*, 185–189. https://doi.org/10.1016/j.gaitpost.2022.05.032

Laakso, H. M., Hietanen, M., Melkas, S., Sibolt, G., Curtze, S., Virta, M., Ylikoski, R., Pohjasvaara, T., Kaste, M., Erkinjuntti, T., & Jokinen, H. (2019). Executive function subdomains are associated with post-stroke functional outcome and permanent institutionalization. *European Journal of Neurology*, *26*(3), 546–552. https://doi.org/10.1111/ene.13854

Maeir, A., Krauss, S., & Katz, N. (2011). Ecological validity of the multiple errands test (MET) on discharge from neurorehabilitation hospital. *OTJR: Occupation, Participation and Health*, *31*(1_suppl), S38–S46. https://doi.org/10.3928/15394492-20101108-07

Mahoney, F., & Barthel, D. W. (1965). Functional evaluation ; the Barthel index. A simple index of the independence useful in scoring improvement in the rehabilitation of the chronically ill. *Maryland State Medical Journal*, *14*, 61–65.

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language*, *21* (2), 169–183. https://doi.org/10.1055/s-2000-7563

Mole, J. A., & Demeyere, N. (2020). The relationship between early post-stroke cognition and longer term activities and participation: A systematic review. *Neuropsychological Rehabilitation*, *30*(2), 346–370. https://doi.org/10.1080/09602011.2018.1464934

Moore, M. J., Vancleef, K., Riddoch, M. J., Gillebert, C. R., & Demeyere, N. (2021). Recovery of visuospatial neglect subtypes and relationship to functional outcome six months after stroke. *Neurorehabilitation and Neural Repair*, *35*(9), 823–835. https://doi.org/10.1177/15459683211032977

Morrison, M. T., Giles, G. M., Ryan, J. D., Baum, C. M., Dromerick, A. W., Polatajko, H. J., & Edwards, D. F. (2013). Multiple Errands Test–Revised (MET–R): A performance-based measure of executive function in people with mild cerebrovascular accident. *American Journal of Occupational Therapy*, *67*(4), 460–468. https://doi.org/10.5014/ajot.2013.007880

Nagamatsu, L. S., Voss, M., Neider, M. B., Gaspar, J. G., Handy, T. C., Kramer, A. F., & Liu-Ambrose, T. Y. L. (2011). Increased cognitive load leads to impaired mobility decisions in seniors at risk for falls. *Psychology and Aging*, *26*(2), 253–259. https://doi.org/10.1037/a0022929

Nalder, E. J., Clark, A. J., Anderson, N. D., & Dawson, D. R. (2017). Clinicians' perceptions of the clinical utility of the multiple errands test for adults with neurological conditions. *Neuropsychological Rehabilitation*, *27*(5), 685–706. https://doi.org/10.1080/09602011.2015.1067628

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nys, G. M. S., van der Worp, H. B., van Zandvoort, M. J. E., de Haan, E. H. F., de Kort, P. L. M., Jansen, B. P. W., & Kappelle, L. J. (2006). Early cognitive impairment predicts long-term depressive symptoms and quality of life after stroke. *Journal of the Neurological Sciences*, *247*(2), 149–156. https://doi.org/10.1016/j.jns.2006.04.005

Ownsworth, T., & Shum, D. (2008). Relationship between executive functions and productivity outcomes following stroke. *Disability and Rehabilitation*, *30*(7), 531–540. https://doi.org/10.1080/09638280701355694

Parikh, R. M., Robinson, R. G., Lipsey, J. R., Starkstein, S. E., Fedoroff, J. P., & Price, T. R. (1990). The impact of poststroke depression on recovery in activities of daily living over a 2-year follow-up. *Archives of Neurology*, *47*(7), 785–789. https://doi.org/10.1001/archneur.1990.00530070083014

Pohjasvaara, T., Leskelä, M., Vataja, R., Kalska, H., Ylikoski, R., Hietanen, M., Leppävuori, A., Kaste, M., & Erkinjuntti, T. (2002). Post-stroke depression, executive dysfunction and functional outcome. *European Journal of Neurology*, *9*(3), 269–275. https://doi.org/10.1046/j.1468-1331.2002.00396.x

Quinn, T. J., Elliott, E., & Langhorne, P. (2018). Cognitive and mood assessment tools for use in stroke. *Stroke*, *49*(2), 483–490. https://doi.org/10.1161/STROKEAHA.117.016994

Quinn, T. J., Langhorne, P., & Stott, D. J. (2011). Barthel index for stroke trials. *Stroke*, *42*(4), 1146–1151. https://doi.org/10.1161/STROKEAHA.110.598540

Rand, D., & Katz, N. (2009). Training multitasking in a virtual supermarket: A novel intervention after stroke. *American Journal of Occupational Therapy*, *63*(5). https://doi.org/10.5014/ajot.63.5.535simm

Rand, D., Katz, N., Shahar, M., Kizony, R., & Weiss, P. L. (2005). The virtual mall: A functional virtual environment for stroke rehabilitation. *Annual Review of Cybertherapy and Telemedicine: A Decade of VR*, *3*, 193–198.

Rand, D., Rukan, S. B.-A., Weiss, P. L., & Katz, N. (2009). Validation of the virtual MET as an assessment tool for executive functions. *Neuropsychological Rehabilitation*, *19*(4), 583–602. https://doi.org/10.1080/09602010802469074

Rankin, J. (1957). Cerebral vascular accidents in patients over the age of 60: II. *Prognosis. Scottish Medical Journal*, *2*(5), 200–215. https://doi.org/10.1177/003693305700200504

Raspelli, S., Pallavicini, F., Carelli, L., Poletti, B., Corra, B., Silani, V., & Riva, G. (2011). Validation of a neuro virtual reality- based version of the multiple errands test for the assessment of executive functions. *Studies in Health Technology and Informatics*, *167*, 92–97.

Robnett, R. H., Hahn, K., & Roland, T. (2021). The development of a multiple errands test for pre/post concussive testing on a college campus: The university multiple errands test-lessons learned.

Roper, A. (2017). *Computer gesture therapy for adults with severe aphasia*. University Lonon. https://openaccess.city.ac.uk/id/eprint/18030/

Rotenberg, S., Ruthralingam, M., Hnatiw, B., Neufeld, K., Yuzwa, K. E., Arbel, I., & Dawson, D. R. (2020). Measurement properties of the multiple errands test: A systematic review. *Archives of Physical Medicine and Rehabilitation*, *101*(9), 1628–1842. https://doi.org/10.1016/j.apmr.2020.01.019

Salter, K. L., Teasell, R. W., Foley, N. C., & Jutai, J. W. (2007). Outcome assessment in randomized controlled trials of stroke rehabilitation. *American Journal of Physical Medicine & Rehabilitation*, *86*(12), 1007–1012. https://doi.org/10.1097/PHM.0b013e3181587b3d

Scarff, S. M., Nalder, E. J., Gullo, H. L., & Fleming, J. (2022). The multiple errands test: A guide for site-specific version development. *Canadian Journal of Occupational Therapy*, 00084174221142184. https://doi.org/10.1177/00084174221142184

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Shallice, T., & Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain : A Journal of Neurology*, *114*(2), 727–741. https://doi.org/10.1093/brain/114.2.727

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Available at SSRN 2160588*. https://doi.org/10.2139/ssrn.2160588

Snyder, H., Miyake, A., & Hankin, B. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers In Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00328

Tian, Y., Thompson, J., Buck, D., & Sonola, L. (2013). *Exploring the system-wide costs of falls in older people in Torbay* (ISBN978 1 909029 15 6; p. 12). The Kings Fund. 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

Tien, A. Y., Spevack, T. V., Jones, D. W., Pearlson, G. D., Schlaepfer, T. E., & Strauss, M. E. (1996). Computerized Wisconsin card sorting test: Comparison with manual administration. *The Kaohsiung Journal of Medical Sciences*, *12*(8), 479–485.

van der Zwaluw, C. S., Valentijn, S. A. M., Nieuwenhuis-Mark, R., Rasquin, S. M. C., & van Heugten, C. M. (2011). Cognitive functioning in the acute phase poststroke: A predictor of discharge destination? *Journal of Stroke and Cerebrovascular Diseases*, *20*(6), 549–555. https://doi.org/10.1016/j.jstrokecerebrovasdis.2010.03.009

Verstraeten, S., Mark, R., & Sitskoorn, M. (2016). Motor and cognitive impairment after stroke: A common bond or a simultaneous deficit. *Stroke Research & Therapy*, *1*(1), 1.

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *The Lancet*, *370* (9596), 1453–1457. https://doi.org/10.1016/S0140-6736(07)61602-X

Webb, S. S., Anders, J., Chiu, E. G., Payne, F., Basting, R., Duta, M. D., & Demeyere, N. (2021). The Oxford Digital Multiple Errands Test (OxMET). *Validation of a Simplified Computer Tablet Based Multiple Errands Test. Neuropsychological Rehabilitation*, https://doi.org/10.1080/09602011.2020.1862679

Webb, S. S., & Demeyere, N. (2021). Multiverse analysis: A method to determine researcher degrees of freedom in test validation. https://doi.org/10.31234/osf.io/nhrwq.

Webb, S. S., & Demeyere, N. (2023). Predictive validity of the Oxford Digital Multiple Errands Test (OxMET) for functional outcomes after stroke. *PsyArXiv*. https://doi.org/10.31234/osf.io/agb3j

Wei, W. E., De Silva, D. A., Chang, H. M., Yao, J., Matchar, D. B., Young, S. H. Y., See, S. J., Lim, G. H., Wong, T. H., & Venketasubramanian, N. (2019). Post-stroke patients with moderate function have the greatest risk of falls: A national cohort study. *BMC Geriatrics*, *19*(1), 373. https://doi.org/10.1186/s12877-019-1377-7