
Multiple Errands Test–Revised (MET–R): A Performance-Based Measure of Executive Function in People With Mild Cerebrovascular Accident

M. Tracy Morrison, Gordon Muir Giles, Jennifer D. Ryan,
Carolyn M. Baum, Alexander W. Dromerick, Helene J. Polatajko,
Dorothy F. Edwards

MeSH TERMS

- activities of daily living
- executive function
- reproducibility of results
- stroke
- task performance and analysis

OBJECTIVE. This article describes a performance-based measure of executive function, the Multiple Errands Test–Revised (MET–R), and examines its ability to discriminate between people with mild cerebrovascular accident (mCVA) and control participants.

METHOD. We compared the MET–R scores and measures of CVA outcome of 25 participants 6 mo post-mCVA and 21 matched control participants.

RESULTS. Participants with mCVA showed no to minimal impairment on measures of executive function at hospital discharge but reported difficulty with community integration at 6 mo. The MET–R discriminated between participants with and without mCVA ($p \leq .002$).

CONCLUSION. The MET–R is a valid and reliable measure of executive functions appropriate for the evaluation of clients with mild executive function deficits who need occupational therapy to fully participate in community living.

M. Tracy Morrison, OTD, is Clinical Scientist, Sister Kenny Rehabilitation Institute, Minneapolis, MN, and Assistant Professor, University of Kansas Medical Center Department of Occupational Therapy, Kansas City; mtracymorrison@gmail.com

Gordon Muir Giles, PhD, OTR/L, FAOTA, is Assistant Professor, Samuel Merritt University, Oakland, CA.

Jennifer D. Ryan, PhD, is Professor, Rotman Research Institute and Departments of Psychology and Psychiatry, University of Toronto, Toronto, Ontario.

Carolyn M. Baum, PhD, OTR/L, is Professor, Washington University School of Medicine, Program in Occupational Therapy, St. Louis, MO.

Alexander W. Dromerick, MD, is Professor of Rehabilitation Medicine and Neurology, National Rehabilitation Hospital and Georgetown University, Washington, DC.

Helene J. Polatajko, PhD, OT Reg. (Ont.), OT(C), is Professor, Department of Occupational Science and Occupational Therapy, University of Toronto, Toronto, Ontario.

Dorothy F. Edwards, PhD, is Professor, Department of Kinesiology–Occupational Therapy, University of Wisconsin–Madison.

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, 460–468. <http://dx.doi.org/10.5014/ajot.2013.007880>

Executive function (EF) refers to a cluster of thinking abilities identified as high-level cognitive processes distinct from memory and attentional capacities (Baddeley, 1986; Lezak, 1982; Norman & Shallice, 1986; Stuss & Alexander, 2000). EF enables people to establish purposeful goals and determine effective strategies that support goal attainment (Tranel, Hathaway-Nepple, & Anderson, 2007). EF involves the orchestration of many complex behaviors, and even mild deficits in EF can result in significant functional deficits and decline in occupational performance (Eslinger, Moore, Anderson, & Grossman, 2011; Foster et al., 2011).

Mild to moderate EF deficits are notoriously difficult to identify with customary neuropsychological or occupational therapy evaluations (Chan, Shum, Touloupoulou, & Chen, 2008). To date, no universally accepted assessments of EF are available (Schmitter-Edgecombe, Parsey, & Cook, 2011; Stuss & Alexander, 2000). Traditional neuropsychological tests fail to identify subtle EF deficits that are serious enough to result in real-world functional decline (Manchester, Priestley, & Jackson, 2004). Similarly, functional assessments commonly present routine tasks that prioritize procedural memory and may fail to capture EF impairments (Miyake et al., 2000). In fact, depending on the degree of task familiarity, people with EF impairments may perform activities of daily living (ADLs) and instrumental activities of daily living (IADLs) without difficulty (Edwards, Hahn, Baum, & Dromerick, 2006).

The full range of EF deficits may best be identified through performance-based tests (PBTs) that use functional tasks but that add rules either constraining how the test taker can approach the task or defining task goals in novel ways (Burgess et al., 2006; Kibby, Schmitter-Edgecombe, & Long, 1998; Shallice & Burgess, 1991). PBTs of EF assess strategies used to accomplish purposeful goals. People use EFs if they perceive a task to be challenging yet within their capacity and if they are motivated to accomplish the task (Lezak, 1982; Schutz & Wanlass, 2009). PBTs may be inherently motivating because test results are easily recognized as relevant to functional abilities in everyday life (Schutz & Wanlass, 2009). Although more ecologically valid than neuropsychological measures, PBTs often lack standardized administration procedures and scoring methods and test reliability and validity (Sadek, Stricker, Adair, & Haaland, 2011). The factors that help PBTs accurately assess EF are not well delineated but include the nature of the environment, the novel and dynamic nature of the test conditions, and—in some PBTs—the high cognitive load.

The contribution of test environments to the challenge presented in the PBTs of EF is unclear. Two distinct types of PBT can be identified on the basis of testing context: laboratory-based and real-world measures. Laboratory-based measures control for environmental variables to maintain test consistency (Giovannetti, Schmidt, Gallo, Sestito, & Libon, 2006). Measures administered in real-world environments require task performance in the face of unpredictable affordances and interpersonal interactions (e.g., in a grocery store; Hamera & Brown, 2000; Hamera, Rempfer, & Brown, 2005). Real-world measures typically mandate little or no assistance from the tester. The scores of real-world measures capture participants' actions throughout a test and address contextual influences contributing to test performance (e.g., noise levels, social demands). Thus, real-world testing is presumed to more adequately capture day-to-day performance limitations in complex life tasks. Prior studies have suggested that PBTs with novel and dynamic test conditions expose participants to heightened challenges, which may increase the sensitivity of such tests to EF deficits among people with mild neurological impairment. Similarly, task simplicity and overlap of tasks with clients' normal and customary routines create additional levels of challenge in inhibitory control and set maintenance (Burgess, 2000; Shallice & Burgess, 1991).

One of the first measures to incorporate these factors into a PBT test, and the most widely researched, is the Multiple Errands Test (MET; Shallice & Burgess, 1991). The MET was developed to measure action-dependent EF among community-dwelling people with mild neurocognitive impairment (Shallice & Burgess, 1991). During the MET,

participants perform multiple everyday tasks in a real-world environment. The simplicity of the MET tasks and the environmental conditions coupled with the rule restrictions create a highly challenging test scenario (Burgess, 2000; Dawson et al., 2009).

Because the MET is administered in real-world environments, local versions should be developed. Several versions of the original MET have been described, including a simplified U.K. hospital version (Knight, Alderman, & Burgess, 2002), a simplified version administered in a U.K. shopping mall (Alderman, Burgess, Knight, & Henman, 2003), a Canadian hospital version (Dawson et al., 2009), and virtual reality versions (Rand, Basha-Abu Rukan, Weiss, & Katz, 2009; Raspelli et al., 2010).

The MET demonstrates promise in detecting EF deficits among people with little or no impairment on neuropsychological tests but who exhibit ongoing real-world difficulties (Shallice & Burgess, 1991). Problems with the MET must be solved, however, before occupational therapists can use it in routine clinical practice. Current scoring paradigms for the MET rely largely on subjective rater impressions, limiting its clinical utility. Additionally, no current MET scoring system discriminates between neurologically healthy people and those with mild brain damage (Alderman et al., 2003; Dawson et al., 2009; Knight et al., 2002). Accordingly, we developed a new version of the MET, the Multiple Errands Test–Revised (MET–R), to provide an objective scoring system and to improve clinical utility. Given that no published studies have examined people with very mild neurological impairment, our goal for the current study was to examine the preliminary discriminant validity of the MET–R among a group with mild impairment. Additionally, we examined both the interrater reliability of the MET–R scoring system and its concurrent validity by comparing findings on the MET–R to those obtained with the Executive Function Performance Test (EFPT; Baum et al., 2008). Finally, we examined the predictive validity of the MET–R by examining the relationship of scores on the MET–R to self-reported community integration after an mCVA.

Method

Research Design

This prospective cohort study included three phases of MET–R development: (1) design; (2) initial assessment of interrater reliability; and (3) determination of discriminant validity, concurrent validity, and clinical utility. The appropriate institutional review board approved data collection procedures. All participants provided informed consent.

Participants

A convenience sample of participants with mCVA were recruited from the acute neurology stroke service of a large university-affiliated tertiary care hospital. We recruited patients with first-time ischemic stroke with scores of ≤ 5 on the National Institutes of Health Stroke Scale (NIHSS; Brott et al., 1989). A stroke neurologist reviewed all cases to verify stroke diagnosis prior to commencement of the study. Patients with a history of prior CVA, depression, dementia, psychosis, or premorbid functional impairment (Barthel Index < 95 ; Modified Rankin Scale > 1) were excluded (Bonita & Beaglehole, 1988; Mahoney & Barthel, 1965). The baseline assessment was completed during the acute hospital stay; in-person evaluations were conducted 6 mo post-mCVA. All participants were discharged directly to their homes, and all were living independently in the community at the time of the 6-mo evaluations. Matched healthy control participants were screened and recruited through the university's Volunteers for Health program.

Instruments

This study used four measures: the NIHSS, the FIM™ (Hamilton, Granger, Sherwin, Zielesny, & Tashman, 1987), the Stroke Impact Scale 2.0 (SIS; Duncan et al., 1999), and the EFPT. The NIHSS assesses stroke-related neurological impairment. Total score interrater reliability is reported as high (intraclass correlation coefficient [ICC] = .80) with good predictive validity ($r = .70$). Scores range from 0 (*no deficit*) to 46 (*severe deficit*).

The FIM (Hamilton et al., 1987) uses 18 items to grade the level of cognitive and physical assistance necessary for function. High internal consistency has been reported in acute stroke ($\alpha = .91$), with excellent concurrent validity with the Barthel Index ($r = .94$). Item scores range from 1 (*completely dependent*) to 7 (*completely independent*), and total scores range from 18 to 126.

The SIS assesses self-reported impact of stroke on eight domains—strength, hand function, ADLs and IADLs, mobility, communication, emotion, memory, and participation—and provides a global stroke recovery index. Reliability is reported to be high (.83 to .90), with good predictive validity with the Short Form-36 Health Survey ($r = .77$; McHorney, Ware, & Raczek, 1993). Domain scores range from 0 to 100; higher scores represent more positive outcomes.

The EFPT measures EF abilities through participants' performance of four test tasks important for independent living: cooking, managing medications, using telephones, and paying bills. Interrater reliability was high (ICC = .91);

discriminant validity was also established ($p < .05$). Examiners use a 0–5 ordinal scoring system according to the level of cueing needed for each task element; scores of 2 or greater indicate impairment. Overall total scores range from 0 to 100, with 0 indicating perfect performance on all functional tasks.

Development of the Multiple Errands Test–Revised

The MET–R was modeled on the original MET and the simplified hospital versions (Knight et al., 2002). In accordance with the original MET, tasks were interleaved, and several tasks could be accomplished at a single location. We developed an administration manual to ensure that all participants received standard instructions and to prevent inadvertent cues from the examiner. We also created an observation form, shown in Figure 1, to record the number of locations visited and the path participants took to complete the assigned tasks, with frequency ratings assigned to rule breaks (see Table 1). *Rule breaks* are identified as an action taken by the participant that violates rules listed on the task list. An example of a rule break is revisiting the same location (e.g., the gift shop), which violates the rule “do not go back into an area you have already been into.”

We created a simple scoring system with the following measures of performance: total time, number of locations visited, number of tasks completed, and total rule breaks. Total rule breaks is the sum of all types of rule breaks and their frequencies. Scoring for total rule breaks is an important difference from prior MET scoring systems and assumes that frequency is more sensitive than type of rule break in discriminating between groups. The MET–R features a new score, *performance efficiency*, which is a ratio of tasks completed to total number of locations visited. A perfect efficiency score is 3.4, calculated by dividing the number of tasks (17) by the number of locations (5) that participants have to visit to accomplish all the tasks. The calculated performance efficiency was normalized to a ratio between 0 and 1; a perfect score is 1.0. The long-term objective was to create an ecologically valid standardized index that would combine two important measures of the MET–R. Our scoring system also departs from prior test versions in that we removed the concepts of inefficiency and interpretation failures, which were too subjective to be scored easily (Dawson et al., 2009; Shallice & Burgess, 1991). Additionally, we removed task failures to avoid counting the same error twice.

Administration Procedures

The MET–R was administered on the main floor of a large hospital. Participants were taken to a central location in the hospital lobby and asked to establish their

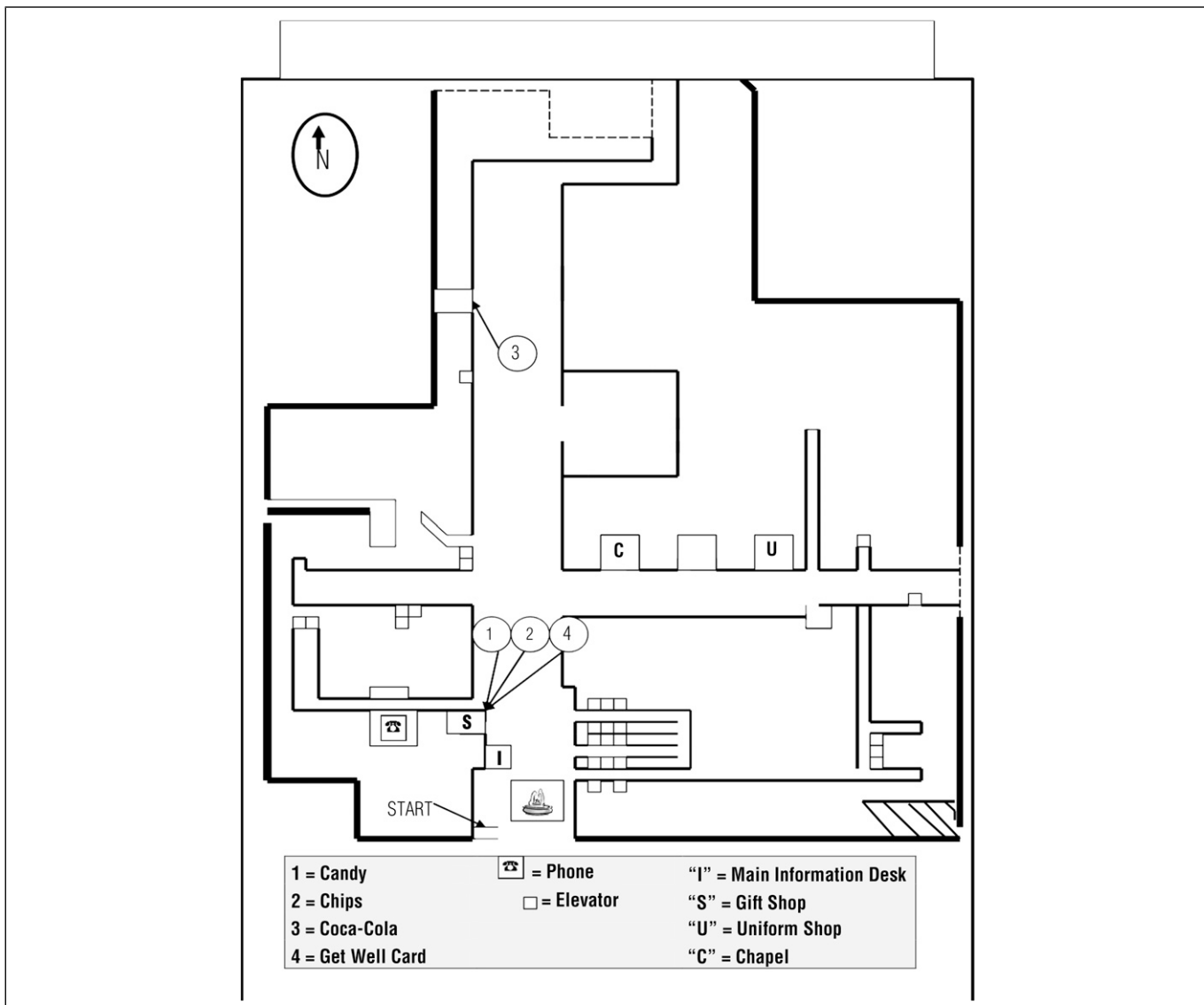


Figure 1. Map used to score the Multiple Errands Test–Revised with required locations marked.

location by examining a map of the main hospital floor and marking an “X” on the map indicating their starting location. Next, participants were asked to read both the MET–R task instructions and the list of rules to the examiner. The examiner gave standardized responses to all inquiries. Participants were given a clipboard, a task list, a map, money, a pen, and a backpack and instructed to self-initiate the test without prompting. During testing, the examiner followed participants through the hospital and used the scoring template to document participant performance. Participants were instructed that without further prompting, they should tell the examiner when they had completed the test. The examiner would have stopped participants who exceeded a 45-min time limit, but none did so. Before initiating the MET–R study, we sequentially recruited 10 participants from the mCVA group who were approaching the date for 6-mo follow-up visits for evaluation of interrater reliability, and their

performance on the MET–R was videotaped. Two raters independently observed the videotapes and scored the MET–R using the standardized observation and scoring protocol. Minor changes were made to the study procedures on the basis of this process. Study data collection commenced after determination of interrater reliability.

Data Collection

Demographic and social information and health history (including acute neurological and treatment data) were collected during the acute hospitalization. The 6-mo mCVA outcome variables were obtained during an in-person visit and included a comprehensive neuropsychological assessment battery, the SIS, the EFPT, and the MET–R. Control participants were evaluated on the same variables minus the SIS. The first author (Morrison) collected the MET–R data while blinded to participant status, acute test data, and the other 6-mo outcome data.

Table 1. Items and Definitions on the MET-R

| MET-R Scoring Item | Definition |
|-----------------------------------|---|
| Total time to complete | Total time elapsed from start to when the participant tells the examiner he or she is finished |
| Total number of locations visited | Total number of locations visited, regardless of whether the locations are related to a task |
| Total number of tasks completed | Number of the 17 tasks assigned that were finished |
| Total number of passes | Number of times the participant visited any location relevant to the test tasks |
| Total number of rule breaks | Number of instructions violated (e.g., if a participant undertook three nonrequisite lexical interactions with the examiner, the calculation would include these three instances) |

Note. MET-R = Multiple Errands Test-Revised.

Statistical Analysis

Descriptive statistics were computed for continuous data and frequency distributions for noncontinuous data. We used χ^2 tests to compare frequency distributions between the mCVA and control groups. We used Student's *t* tests to compare mCVA participants with control participants regarding age, education, MET-R scores, and EFPT total score. Interrater reliability was assessed via intraclass correlation coefficients. Pearson correlation coefficients were computed to compare scores on the MET-R with scores on the EFPT (Portney & Watkins, 2009). We used *p* values <.05 corrected to <.01 based on Bonferroni corrections for multiple comparisons as the criteria for significance. IBM SPSS Version 20 (IBM Corporation, Armonk, NY) was used for statistical analysis.

Results

The baseline assessment was completed with 50 patients during the acute hospital stay; in-person evaluations were conducted 6 mo post-mCVA. Two of the 50 patients had subsequent strokes after discharge and before the 6-mo follow-up, 1 patient was lost to follow-up, and 2 declined to participate. Of the 45 participants assessed at 6 mo, 25 were recruited sequentially for participation in the current study. Twenty-one matched healthy control participants were included in this study.

Table 2 presents the participants' demographic and clinical characteristics. Control and mCVA participants did not differ significantly with respect to age, sex, race, or years of education.

The NIHSS mean score for the mCVA participants at the time of admission to the acute stroke unit was 3.35 (standard deviation [*SD*] = 1.50). Of the mCVA participants, 88% had NIHSS scores between 0 and 3, and none exceeded a score of 5 at the time of admission (Table 2). Mean FIM score at hospital discharge was 118 (*SD* = 4.32).

The mean NIHSS score at the 6-mo follow-up was 2.42 (*SD* = 2.10), and all FIM scores were at maximum (126), indicating no impairment. Thus, these participants were functionally independent and had minimal levels of

residual neurological impairment. The MET-R and the EFPT were also administered during this study visit.

Self-reported CVA outcome using the SIS was also assessed at this study visit. The scores on this measure provide evidence of an mCVA's impact on participants' occupational performance and community participation. Despite minimal residual neurological impairment and functional independence, these self-report scores suggest that the study participants had not achieved complete recovery. The SIS scores reflect a number of areas of mCVA-related difficulties. The mean global recovery rating was only 68.60 (*SD* = 21.60) out of a possible score of 100. Scores on the mCVA Participation, Memory, and Emotion subscales were much lower than the near-maximum scores on the ADL and Mobility subscales (see Table 3).

Reliability

In the first phase of the study, before the assessment of the participants, we established interrater reliability on the MET-R. We compared the two independent raters' scores for all MET-R sections (i.e., tasks completed, total rule breaks, total locations, number of passes, and total time) and found the scores to be identical, resulting in ICCs of 1 for all test sections.

Validity

Examination of the mean scores for each MET-R section (see Table 3) shows that although the control group performed better than the mCVA group on all sections, several of the control participants had difficulty with some test items. For example, the control participants on average completed only about 15 of the 17 tasks, and some broke rules during testing. Differences between the control participants and the mCVA group were significant for the majority of MET-R component scores, including total tasks completed (*p* < .001), rule breaks (*p* < .001), and performance efficiency (*p* < .002; see Table 3). These results support the discriminant validity of the MET-R.

EFPT data were available for 20 of the 21 control participants. The mean EFPT score was 2.20 (*SD* = 1.10) for the control participants and 4.09 (*SD* = 2.25) for the

Table 2. Demographic and Baseline Clinical Characteristics of Study Groups

| Variable | Control Group (<i>n</i> = 21) | mCVA Group (<i>n</i> = 25) | <i>p</i> |
|-----------------------------|--------------------------------|-----------------------------|----------|
| Age, yr (<i>SD</i>) | 59.9 ± 15.50 | 60.0 ± 10.80 | .98 |
| Education, yr (<i>SD</i>) | 14.8 ± 2.50 | 13.4 ± 2.70 | .07 |
| Admission NIHSS score | NA | 3.35 ± 1.50 | NA |
| Sex—female, <i>n</i> (%) | 15 (71) | 15 (60) | .62 |
| Race | | | |
| White, <i>n</i> (%) | 17 (81) | 15 (60) | .20 |
| Black, <i>n</i> (%) | 4 (19) | 10 (40) | |

Note. mCVA = mild cerebrovascular accident; NA = not applicable; NIHSS = National Institutes of Health Stroke Scale; *SD* = standard deviation.

mCVA participants. No significant differences were found between the groups on the EFPT. The correlation between the MET-R task completion score and the EFPT total score was $r = -.55$. This moderate correlation was expected given the differences in examiner support, task demands, and level of structure in the testing environment. Thus, the MET-R demonstrated concurrent validity with the EFPT, which is a validated measure of a person's EF after a CVA.

Discussion

Our findings support the reliability and validity of the MET-R as a PBT of EF for people with mCVA. Our goal was to determine whether a real-world PBT of EF, the MET-R, could distinguish between people with mCVA and matched neurologically healthy control participants. Our interest in the MET developed because of the clinical difficulty in identifying people who present without overt deficits in memory, attention, or motor skills but who experience functional deficits attributable to EF impairments. The MET-R's refinements enhance its ease of use and its relevance in clinical practice. This

study demonstrates that the new MET-R scoring method is sensitive to differences between a group of people with mCVA and a group of healthy people. The total number of rule breaks was found to be highly sensitive to post-mCVA EF deficits such that approximately 88% of the control group had fewer total rule breaks than the average person in the mCVA group. Performance efficiency also reflected between-group differences; approximately 69% of the control group performed more efficiently than the average person in the mCVA group. Thus, the MET-R remains consistent with the intent of the test's original authors in that it emphasizes real-world task performance deficits in people who, although they appear unimpaired on routine testing, experience functional disability in real-world environments (Shallice & Burgess, 1991).

Various methods can help adjust the level of challenge presented in PBTs of EF. Prior studies have suggested that task novelty increases test-challenge levels and, therefore, facilitates EF assessments (Park, et al., 2012). Nonetheless, the majority of PBTs address familiar, everyday life tasks rather than novel ones (Baum et al., 2008). Increased challenge (cognitive load) can be

Table 3. Six-Month Outcome Scores

| Variable | Control Group (<i>n</i> = 21), <i>M</i> ± <i>SD</i> | mCVA Group (<i>n</i> = 25), <i>M</i> ± <i>SD</i> | <i>p</i> |
|-----------------------------|--|---|----------|
| MET-R scores | | | |
| Performance efficiency | 0.50 ± 0.20 | 0.30 ± 0.20 | .002 |
| Total tasks completed | 14.70 ± 2.10 | 11.0 ± 0.20 | .001 |
| Total number of rule breaks | 3.80 ± 4.20 | 8.30 ± 4.00 | .001 |
| Total locations | 9.00 ± 2.70 | 11.00 ± 4.30 | .08 |
| Total time | 18.00 ± 6.16 | 20.13 ± 7.36 | .25 |
| EFPT total score | 2.20 ± 1.10 | 4.09 ± 2.25 | .40 |
| SIS scores | | | |
| Global recovery | NA | 68.60 ± 21.60 | NA |
| Participation | NA | 73.80 ± 24.10 | NA |
| Memory | NA | 77.10 ± 17.50 | NA |
| Emotion | NA | 76.00 ± 21.60 | NA |
| Activities of daily living | NA | 92.40 ± 9.70 | NA |
| Mobility | NA | 86.00 ± 11.40 | NA |

Note. EFPT = Executive Function Performance Test; *M* = mean; mCVA = mild cerebrovascular accident; MET-R = Multiple Errands Test-Revised; NA = not applicable; *SD* = standard deviation; SIS = Stroke Impact Scale.

created through the use of multiple overlapping task demands, the addition of time pressure, or subtle obstacles within the functional task. Increased challenge can also be created through novel rules acting as constraints on task performance. Novel rules may be particularly challenging in the context of overlearned functional tasks because the rules require people to adjust familiar behaviors and routines or engage in set shifting to achieve test goals (Schutz & Wanlass, 2009). We often observed participants introduce personally valued but test-irrelevant considerations into their task performance, leading to MET-R errors. Our clinical observations suggest that many clients have similar problems in their daily lives.

Increased task challenge may also be available through the use of multiple and interleaved tasks versus single tasks. In multiple-task scenarios, clients must monitor their progress to successfully perform multiple tasks at the same time. In some circumstances, for a client to undertake certain tasks, he or she must refrain from completing other tasks, and so successful MET-R performance is highly dependent on both planning and response inhibition.

We hypothesized that the MET-R's suitability for relatively complex, dynamic contexts would translate into increased sensitivity to subtle executive deficits. We found that both the MET-R and the EFPT detected deficits in our study participants. The EFPT identified mild deficits in mCVA participants who, on average, required six cues to complete EFPT tasks. The MET-R identified sometimes quite dramatic performance deficits among the mCVA population and therefore, as expected, may be more sensitive than the EFPT to EF deficits among people with mild impairment. Quantifying this difference, however, requires further research and a larger study population.

This study's mCVA participants had minimal or absent deficits on standard CVA rehabilitation measures (FIM, NIHSS), and consequently no mCVA participants received rehabilitation services beyond their brief (<72 hr) acute hospital stay. In general, our mCVA participants, when assessed 6 mo postinjury, were aware of their deficits to the extent that they would self-impose restrictions on their own daily activities. The test environment of the MET-R was a consistently busy hospital lobby with hospital staff and patients roaming throughout the area. During test debriefing, many mCVA participants stated that they found the hospital context intimidating and that they had similar feelings when out in their own communities. After discharge from the hospital to home, this cohort did not return to pre-mCVA work or participation levels despite being, on

average, young and highly educated. These findings echo the findings of a similar study investigating an mCVA population (O'Brien & Wolf, 2010).

The MET-R differs from some EF-oriented PBTs in that it involves multiple, combined, interleaved tasks that, on their own, appear routine. Additionally, the MET-R requires participants to interact with—and, more specifically, to obtain information from—strangers, a task requiring social knowledge and personal confidence. The unpredictable contexts associated with the MET-R could derail task performance and, therefore, could present heightened challenges during testing. Because of these factors and the MET-R's rule constraints, the performance of combined tasks most likely requires EF ability. Earlier MET versions lacked a scoring system that is reliable and that can be used consistently across local MET variants. By contrast, in the MET-R, errors are readily quantifiable and are free of the examiner's subjective inferences about the participant's behavioral failures. Performance-based scoring paradigms must be reliable and provide a method for reporting performance differences among individuals and patient groups. We have demonstrated that the MET-R's scoring method is highly reliable. In fact, when we compared the two raters' scores, all items had perfect ICCs.

To our knowledge, this is the first study to establish a MET variant possessing sensitivity to real-world EF impairments among a homogeneous group of people with first-time mCVA and mild impairment. Prior MET studies distinguished between a neurologically healthy population and a target population with brain injury of mixed severity (i.e., moderate to severe injury; Dawson et al., 2009; Knight et al., 2002).

The important effects of mild brain injury on the execution of EFs and other valued tasks is increasingly being recognized. Moreover, growing evidence has shown that demographic changes are influencing today's population of people with CVA: Fewer people are dying of CVA, and in the aggregate, CVAs are less severe and occur at younger ages than previously (Wolf, Baum, & Connor, 2009). Consequently, clients are expected to resume community life. The clinical practices of occupational therapists should match the population needing services (Wolf et al., 2009). Measures can help identify clients who, although experiencing EF deficits, are overlooked by current clinical assessments.

The MET-R is a complex test to administer, and for many clients, a test of lesser complexity can satisfactorily reveal EF impairment serious enough to impair real-world performance. On the basis of our experience with the MET-R, we suggest that the MET-R be used with clients in an acute setting who may not fully understand the

nature of their impairments, who have high task demands in their daily lives, and who are not demonstrating deficits on traditional motor- and language-oriented assessments. This is a population that, despite its need for services, does not currently meet routine rehabilitation criteria and, thus, runs the risk of being overlooked by occupational therapists. The MET-R could function as a final assessment for those who are not otherwise showing impairment and could help clients who are returning to their daily life activities gain a better understanding of the types of difficulties that might surface during this recovery process. Additionally, the MET-R might be useful in assessing any person who, on a routine clinical follow-up after an mCVA, reports challenges with the execution of complex tasks.

Limitations and Future Research

Many of the limitations in this study relate to its small sample size. The new scoring system was effective and distinguished the matched control group from the mCVA group; however, it was not possible either to generate cutoff scores for neurologically healthy participants versus mCVA participants or to establish normative values. We are currently engaged in a larger study with this aim. We found high interrater reliability but did not assess test-retest reliability using repeated administration of the MET-R. Previous research reported that other MET versions had good test-retest reliability with a population whose brain injuries were of mixed severity (Knight et al., 2002). We doubt that the MET-R's test-retest reliability will be as high with this study's mCVA population as with other, more cognitively impaired populations. Of more importance is the stability of the new scoring methods, particularly regarding performance efficiency when used in other settings or with other MET variants. We are currently investigating this issue.

Implications for Occupational Therapy Practice

The results of this study have the following implications for occupational therapy practice:

- Performance-based evaluation of EF using real-world activities is an important area of study for occupational therapists.
- The MET-R is highly sensitive in detecting EF deficits in a group of people with mCVA.
- The MET-R scoring simplifications and new efficiency score improve ease of use.
- The MET-R identifies EF deficits missed by standard paper-and-pencil and office-based measures.

- Occupational therapists have a crucial role in supporting clients' recovery from mCVA, and the MET-R may strengthen occupational therapists' ability to identify clients with EF deficits who need occupational therapy to successfully reintegrate into community living. ▲

Acknowledgments

We acknowledge the OTD graduate students of the Program in Occupational Therapy, Washington University School of Medicine, for assistance with data collection; Marsha Finkelstein, Sister Kenney Research Institute, for assistance with data analysis; and the James S. McDonnell Foundation for their support of this project (9832CRHQUA11).

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*, 31–44. <http://dx.doi.org/10.1017/S1355617703910046>
- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Clarendon Press.
- Baum, C. M., Connor, L. T., Morrison, T., Hahn, M., Dromerick, A. W., & Edwards, D. F. (2008). Reliability, validity, and clinical utility of the Executive Function Performance Test: A measure of executive function in a sample of people with stroke. *American Journal of Occupational Therapy, 62*, 446–455. <http://dx.doi.org/10.5014/ajot.62.4.446>
- Bonita, R., & Beaglehole, R. (1988). Recovery of motor function after stroke. *Stroke, 19*, 1497–1500. <http://dx.doi.org/10.1161/01.STR.19.12.1497>
- Brott, T., Adams, H. P., Jr., Olinger, C. P., Marler, J. R., Barsan, W. G., Biller, J., . . . Hertzberg, V. (1989). Measurements of acute cerebral infarction: A clinical examination scale. *Stroke, 20*, 864–870. <http://dx.doi.org/10.1161/01.STR.20.7.864>
- Burgess, P. W. (2000). Strategy application disorder: The role of the frontal lobes in human multitasking. *Psychological Research, 63*, 279–288. <http://dx.doi.org/10.1007/s004269900006>
- Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. M., Dawson, D. R., . . . Channon, S. (2006). The case for the development and use of “ecologically valid” measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society, 12*, 194–209. <http://dx.doi.org/10.1017/S1355617706060310>
- Chan, R. C. K., Shum, D., Touloupoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology, 23*, 201–216. <http://dx.doi.org/10.1016/j.acn.2007.08.010>
- Dawson, D. R., Anderson, N. D., Burgess, P., Cooper, E., Krpan, K. M., & Stuss, D. T. (2009). Further development of the Multiple Errands Test: Standardized scoring, reliability, and ecological validity for the Baycrest version. *Archives of Physical*

- Medicine and Rehabilitation*, 90(Suppl.), S41–S51. <http://dx.doi.org/10.1016/j.apmr.2009.07.012>
- Duncan, P. W., Wallace, D., Lai, S. M., Johnson, D., Embretson, S., & Laster, L. J. (1999). The Stroke Impact Scale version 2.0: Evaluation of reliability, validity, and sensitivity to change. *Stroke*, 30, 2131–2140. <http://dx.doi.org/10.1161/01.STR.30.10.2131>
- Edwards, D. F., Hahn, M., Baum, C. M., & Dromerick, A. W. (2006). The impact of mild stroke on meaningful activity and life satisfaction. *Journal of Stroke and Cerebrovascular Diseases*, 15, 151–157. <http://dx.doi.org/10.1016/j.jstrokecerebrovasdis.2006.04.001>
- Eslinger, P. J., Moore, P., Anderson, C., & Grossman, M. (2011). Social cognition, executive functioning, and neuroimaging correlates of empathic deficits in frontotemporal dementia. *Journal of Neuropsychiatry and Clinical Neurosciences*, 23, 74–82. <http://dx.doi.org/10.1176/appi.neuropsych.23.1.74>
- Foster, E. R., Cunnane, K. B., Edwards, D. F., Morrison, M. T., Ewald, G. A., Geltman, E. M., & Zazulia, A. R. (2011). Executive dysfunction and depressive symptoms associated with reduced participation of people with severe congestive heart failure. *American Journal of Occupational Therapy*, 65, 306–313. <http://dx.doi.org/10.5014/ajot.2011.000588>
- Giovannetti, T., Schmidt, K. S., Gallo, J. L., Sestito, N., & Libon, D. J. (2006). Everyday action in dementia: Evidence for differential deficits in Alzheimer's disease versus subcortical vascular dementia. *Journal of the International Neuropsychological Society*, 12, 45–53.
- Hamera, E., & Brown, C. E. (2000). Developing a context-based performance measure for persons with schizophrenia: The Test of Grocery Shopping Skills. *American Journal of Occupational Therapy*, 54, 20–25. <http://dx.doi.org/10.5014/ajot.54.1.20>
- Hamera, E., Rempfer, M., & Brown, C. E. (2005). Performance in the “real world”: Update on Test of Grocery Shopping Skills (TOGSS). *Schizophrenia Research*, 78, 111–112, author reply 113–114. <http://dx.doi.org/10.1016/j.schres.2005.04.019>
- Hamilton, B. B., Granger, C. V., Sherwin, F. S., Zielezny, M., & Tashman, J. S. (1987). A uniform national data system for medical rehabilitation. In M. J. Fuhrer (Ed.), *Rehabilitation outcomes: Analysis and measurement* (pp. 137–147). Baltimore: Brookes.
- Kibby, M. Y., Schmitter-Edgecombe, M., & Long, C. J. (1998). Ecological validity of neuropsychological tests: Focus on the California Verbal Learning Test and the Wisconsin Card Sorting Test. *Archives of Clinical Neuropsychology*, 13, 523–534.
- 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, 231–255. <http://dx.doi.org/10.1080/09602010244000039>
- Lezak, M. D. (1982). The problems of assessing executive functions. *International Journal of Psychology*, 17, 281–297. <http://dx.doi.org/10.1080/00207598208247445>
- Mahoney, F. I., & Barthel, D. W. (1965). Functional evaluation: The Barthel Index. *Maryland State Medical Journal*, 14, 61–65.
- Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury*, 18, 1067–1081.
- McHorney, C. A., Ware, J. E., Jr., & Raczek, A. E. (1993). The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care*, 31, 247–263.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100. <http://dx.doi.org/10.1006/cogp.1999.0734>
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York: Plenum Press.
- O'Brien, A. N., & Wolf, T. J. (2010). Determining work outcomes in mild to moderate stroke survivors. *Work*, 36, 441–447.
- Park, N. W., Lombardi, S., Gold, D. A., Tarita-Nistor, L., Gravely, M., Roy, E. A., & Black, S. E. (2012). Effects of familiarity and cognitive function on naturalistic action performance. *Neuropsychology*, 26, 224–237. <http://dx.doi.org/10.1037/a0026324>
- Portney, L. G., & Watkins, M. P. (2009). *Foundations for clinical research: Applications to practice* (3rd ed.). Upper Saddle River, NJ: Pearson Education.
- Rand, D., Basha-Abu Rukan, S., Weiss, P. L., & Katz, N. (2009). Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychological Rehabilitation*, 19, 583–602. <http://dx.doi.org/10.1080/09602010802469074>
- Raspelli, S., Carelli, L., Morganti, F., Poletti, B., Corra, B., Silani, V., & Riva, G. (2010). Implementation of the multiple errands test in a NeuroVR-supermarket: A possible approach. *Studies in Health Technology and Informatics*, 154, 115–119.
- Sadek, J. R., Stricker, N., Adair, J. C., & Haaland, K. Y. (2011). Performance-based everyday functioning after stroke: Relationship with IADL questionnaire and neurocognitive performance. *Journal of the International Neuropsychological Society*, 17, 832–840. <http://dx.doi.org/10.1017/S1355617711000841>
- Schmitter-Edgecombe, M., Parsey, C., & Cook, D. J. (2011). Cognitive correlates of functional performance in older adults: Comparison of self-report, direct observation, and performance-based measures. *Journal of the International Neuropsychological Society*, 17, 853–864. <http://dx.doi.org/10.1017/S1355617711000865>
- Schutz, L. E., & Wanlass, R. L. (2009). Interdisciplinary assessment strategies for capturing the elusive executive. *American Journal of Physical Medicine and Rehabilitation*, 88, 419–422. <http://dx.doi.org/10.1097/PHM.0b013e3181a0e2d3>
- Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, 114, 727–741. <http://dx.doi.org/10.1093/brain/114.2.727>
- Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research*, 63, 289–298. <http://dx.doi.org/10.1007/s004269900007>
- Tranel, D., Hathaway-Neppe, J., & Anderson, S. W. (2007). Impaired behavior on real-world tasks following damage to the ventromedial prefrontal cortex. *Journal of Clinical and Experimental Neuropsychology*, 29, 319–332. <http://dx.doi.org/10.1080/13803390600701376>
- Wolf, T. J., Baum, C. M., & Connor, L. T. (2009). Changing face of stroke: Implications for occupational therapy practice. *American Journal of Occupational Therapy*, 63, 621–625. <http://dx.doi.org/10.5014/ajot.63.5.621>