5-2018

A guide to the Multiple Errands Test

Carly Jessup
University of Tennessee at Chattanooga, qyn453@mocs.utc.edu

A Guide to the Multiple Errands Test

Carly Rose Jessup

Departmental Honors Thesis
The University of Tennessee at Chattanooga
Psychology Department

Examination Date: March 27, 2018

Amanda Clark
Assistant Professor of Psychology
Thesis Director

Jill Shelton
Assistant Professor of Psychology
Department Examiner

David Ferrier
Assistant Professor of Psychology
Department Examiner

Table of Contents

Abstract

Executive function plays an important role in everyday life skills such as planning, organizing, controlling and sustaining attention, and socializing. Until recently, executive function was measured almost exclusively using laboratory-based assessments. However, many participants would score satisfactorily on those laboratory assessments, while having significant difficulty with activities of daily living. A potential solution to this issue is the Multiple Errands Test (MET), which is an ecologically valid assessment that can help predict performance of everyday life skills. The MET is a flexible assessment that can be modified in order to fit the needs of the setting and patient. Indeed, many versions have been examined such as the MET-Hospital Version, the Baycrest MET and Baycrest MET Revised, as well as a Virtual MET. This thesis describes how the MET is used in various ways: as a baseline assessment to determine the degree of executive dysfunction, as a tool to teach patients strategies to use to improve their performance, and as an indicator of executive function improvement following interventions. Future research could focus on improving some of the limitations of the MET, specifically how to best simplify scoring while still capturing inefficient behaviors.

Executive Function Symptoms and Assessment

Interest in the brain's frontal lobes began with the famous study of Phineas Gage, a railroad construction worker who, in 1840, had an iron rod driven through his head, destroying his left frontal lobe. After this injury, people began to notice alarming changes in Gage's behavior. He experienced difficulty in carrying out plans; he appeared to have lost control of his behavior, being rude without regarding the feelings of others. More recently, a study in 1985 by Eslinger and Damasio described the case of an accountant who had a large tumor removed. Six years after this operation, the accountant's day-to-day life was troublesome, despite having an IQ over 130 and performing within normal limits on traditional neuropsychological assessments. He was fired from numerous jobs, went bankrupt, and had two divorces in two years. He struggled greatly with making decisions. Simple tasks would take hours to complete. This accountant is a prime example of how seemingly obvious executive dysfunction can go unnoticed due to normal performance on traditional neuropsychological assessments.

The frontal lobes of the brain are vital for supporting executive functions, which are important for governing behavior, strategizing, planning, self-regulating, performing complex activities, organizing, controlling and sustaining attention, and socializing (Gillen, 2009). While various parts of the brain work together and no single part of the brain does everything by itself, the frontal lobes are considered the most important for executive function. Executive dysfunction can result in difficulty in starting and stopping activities, shifting mental and behavioral sets, learning new tasks, and staying focused (Shallice & Burgess, 1991). While people with executive dysfunction usually retain their long-term memory, their ability to efficiently use their knowledge and memories can be impaired. Even a simple task such as keeping track of time can become very challenging (Goldstein & Naglieri, 2014). Individuals with executive dysfunction

typically struggle with instrumental activities of daily living like medication adherence, financial management, and safe use of transportation (Cahn-Weiner, Boyle, & Malloy, 2002). Consequently, executive dysfunction can be life changing.

Executive dysfunction can result due to brain damage from a stroke or traumatic brain injury; it can also be an indicator of dementia (Torralva et al., 2012). Detection of executive dysfunction is valuable for therapeutic intervention as therapists can determine the extent of one's impairment and create an appropriate treatment plan accordingly. For these reasons, determining one's executive ability after frontal lobe damage or deterioration is essential.

Executive function assessment has not progressed as quickly as assessment of other cognitive functions, perhaps due to its unusual history. In fact, until approximately a hundred years ago, many 'experts' did not believe that the frontal lobes even had any function at all. However, Penfield and Evans (1935) (as cited in Burgess et al., 2006) observed a patient who had suffered from frontal lobe damage. This patient was unable to cook a meal as he struggled to complete the required steps involved. Despite this evidence suggesting impairments in everyday life tasks, the early literature on detection of executive dysfunction focused on using methods that were being used to measure other kinds of cognitive impairments, laboratory-based neuropsychological assessments and patient self-reports. Patient self-reports were largely misleading though because a common symptom of executive function is lacking insight. This means that individuals with executive dysfunction may not be aware of how significant their impairments are. Indeed, a study in 2004 by Malec found that individuals with acquired brain injury typically rated themselves as less impaired than their therapists and significant others who rated them. Thus, self-report of symptoms is often unreliable for those suffering from executive

dysfunction. For this reason, laboratory-based neuropsychological assessments have been favored.

One of the most common laboratory assessments of executive function is the Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948). The WCST consists of four key cards and 128 response cards where the response cards vary in the color, form, and number of shapes depicted. Participants then try to match each response card to a key card based on either the number of items, color, or form. They discover the correct classification principle by trial and error and examiner feedback. Participants continue to match according to the principle they correctly discovered. After ten consecutive correct matches, the classification principle changes without warning, and the participant must determine the new classification principle. The WCST continues until all cards have been sorted or a maximum of six correct sorting criteria have been completed. Though the WCST is not timed, performance is scored based on total errors, total categories achieved, frequency of perseverative and nonperseverative errors, failure to maintain the set of responses, percentage of conceptual-level responses, and number of trials to complete the first category (Jodzio & Biechowska, 2010). The WCST was originally created in 1948 as an assessment of abstract reasoning, concept formation, and response strategies to changing contextual contingencies (Eling, Derckx, & Maes, 2008) but it was not until the 1960s that Milner introduced the WCST as an assessment of frontal lobe function in patients with brain lesions. Previous studies have found that individuals with frontal lobe lesions have considerable difficulty with the aspect of perseveration, the continuation of repetition of a response after task requirements have been changed. This is related to issues of executive dysfunction where it is typical for one to have difficulty adapting to change, especially in unplanned circumstances.

Other examples of laboratory-based executive function assessments include the Tower of London test (Shallice, 1982), which measures problem-solving, the Trail-Making test (Army Individual Test Battery, 1994), which measures processing speed and task switching, the Stroop Test (Golden, 1978), which measures task switching and inhibition, and the Controlled Oral Word Association Test (Benton, Hamsher, & Sivan, 1983), which measures verbal fluency. Oddly enough, many of these tests were not actually originally designed for the purpose of testing executive function. Burgess and colleagues (2006) report that Berg developed the WCST in 1948 primarily to assess concept formation. The Tower of London test was developed to test planning and problem-solving. The Stroop Task, used to test processing speed, was created in 1935, but was not used for the purpose of frontal lobe research until 1974. Despite this fact, researchers have published many articles attempting to use these measures to identify executive impairment in individuals with brain abnormalities. A recurring issue with using these laboratory-based assessments is that some individuals with executive impairment will score normally, while experiencing considerable difficulty in activities of daily life (Cahn-Weiner et al., 2002; Cahn-Weiner, Malloy, Boyle, Marran, & Salloway, 2000; Shallice & Burgess, 1991).

One possible explanation for this discrepancy between performance on laboratory assessments and naturalistic assessments is that in laboratory assessments the examiner can function as the patient's executive function (Torralva et al., 2012). These assessments are therefore quite structured with very little room for interference, which is unlike the real world. While patients certainly should not be set up to fail, they should be set up to perform at a level that is representative of their performance level in the real world. Unstructured, naturalistic, assessments require the patient to choose an approach themselves when there are many different possible ways to act.

In addition to the structured environment in laboratory settings, planning, multi-tasking, or problem solving are typically accomplished through assessments instead of tasks reflecting real life tasks (Rand et al., 2009). This is problematic because there are very few situations in real life where one would have the time and resources to work out a problem with pen and paper. For example, when going to the grocery store, a person may use pen and paper to plan what to buy and even group items according to location in the store, but they must then physically go to the store and carry out their plans. The plans are ineffective if the person forgets the list at home or fails to note the grouping of the groceries on the list. Similarly, executive function cannot always be successfully measured by laboratory tests due to the structured testing environment and small sample of actions that can be observed (Chaytor, Schmitter-Edgecombe, & Burr, 2006). Shallice and Burgess (1991) argue that frontal lobe damage affects the ability to plan and multi-task, something that many language, perception, retrospective memory, and executive function tests fail to measure. In order to assess these abilities, the assessment must require multitasking without feedback from the experimenter.

A common issue with laboratory assessments is that they feature situations that patients would not actually encounter in real life. Typically, one completes an executive assessment to determine one's executive capabilities in the real world, but laboratory task performance often does not correlate with everyday life performance. In other words, it lacks ecological validity. While Chaytor and colleagues (2006) found that a group of executive function assessments (Trail Making Test, WCST, Stroop, and Controlled Oral Word Association Test) accounted for 18 - 20% of the variance in everyday executive ability, there is still a large percentage of variance unaccounted for.

Questionnaires such as the Dysexecutive Questionnaire (DEX; Wilson et al., 1996) and the Brock Adaptive Functioning Questionnaire (BAFQ; Dywan & Segalowitz, 1996) have been beneficial in measuring executive function in an ecologically valid form. The DEX measures how often a patient makes 20 different executive function mistakes. The BAFQ measures how often a patient makes 68 different executive function errors related to planning, initiation, flexibility, excess caution, attention, memory, arousal level, emotionality, impulsivity, aggressiveness, social monitoring, and empathy. However, as mentioned above there are drawbacks to these self-report methods. For one, it is preferred that a caregiver or significant other, complete the questionnaires, as they are typically more accurate, particularly if the caregiver knows the patient and their situation well. However, if the person completing the assessment is only around the person for doctor's appointments and does not spend much time with them in the community, their ratings may not be very beneficial.

In summary, laboratory-based executive function assessments may lack representativeness in that the test situations are dissimilar to those that are encountered in real life. Laboratory assessments may also lack generalizability because performance in the structured laboratory may not predict performance in real life situations (Burgess et al., 2006). While traditional laboratory assessments certainly have their merits, they are often unable to capture the impact of executive dysfunction in a naturalistic environment.

The Multiple Errands Test

The Multiple Errands Test (MET) is an assessment that could address the issue of executive function measures lacking generalizability and representativeness. The MET assesses executive function in a real-life environment such as a shopping mall or grocery store. The

assessment is especially unique because it allows for patients to encounter unexpected circumstances, just like those that could easily happen when actually shopping for groceries or running errands. While several versions of the MET exist, all versions feature the same main components: a list of approximately 10 tasks to complete (e.g. purchase a Sprite, find out what time the store closes on Saturday, call the given number and tell that person what day of the week it is), and a list of approximately six to eight rules to follow (e.g. spend as little money as possible, take as little time to complete the exercise without rushing excessively). Even with these commonalities of tasks and rules, the MET allows for unforeseen, unpredictable events to occur. This significantly contributes to the MET's ability to predict performance outside of the hospital or treatment facility (Maeir et al., 2011); while patients can plan their behaviors to execute the tasks, they cannot prepare for unexpected events in the unstructured task environment. For example, the MET can include an "interruption task" which is an additional task that interrupts the participant while he/she is already completing the assessment. This could be a task such as the participant purchasing a copy of the local newspaper. Another example of an interruption is if the participant runs into a friend while completing the assessment. The participant would not be expecting to see his/her friend and thus, the participant must act in a manner that is sensible considering he/she is completing an assessment. While it would be socially unusual to not speak to your friend, it would also be strategically harmful to speak to the friend for a long amount of time as it would negatively affect test performance.

The original MET, created in Shallice and Burgess (1991), involved closely following three people who had sustained frontal lobe injuries. Case One had a skull fracture that had required multiple surgeries (Shallice & Burgess, 1991). He is described as having extreme difficulty keeping his mind on task, organizing social events, and keeping tidy (cleaning,

laundry, and shopping were all done for him). Case Two suffered from a right frontal depressed skull fracture and an intracerebral hematoma. After surgery to address the hematoma, he attempted to return to his previous job but was not successful. He applied for other jobs but was repeatedly dismissed. He had poor hygiene, did not help with chores around the house, was irresponsible with money, lacked planning skills (such as making a list for the grocery store), and could not carry out tasks when they were planned for him (did not get everything on the shopping list that his wife made him). Case Three fractured her skull and suffered from damage to her frontal and temporal lobes. While she was employed (had the same job for twenty-five years) and lived alone, she did not participate in any new activities or activities that were not required. She reportedly never planned anything and was untidy.

Shallice and Burgess (1991) tested their three case participants with their Multiple Errands Test, intelligence tasks (e.g. Wechsler Adult Intelligence Scale), memory tests (Auditory Verbal Learning, Complex Figure Recall), and "frontal lobe" tasks (Stroop test, Tower of London, a modified WCST), and results were varied. Case One performed within normal range on nearly all of the assessments. Case Two performed satisfactorily on frontal lobe tasks, but did not perform satisfactorily on all of the memory tests. Case Three had mixed results on frontal lobe tasks and memory tests. Interestingly, despite the inconsistency in performance on the laboratory-based neuropsychological assessments, all three case studies demonstrated considerable impairment on the MET. This demonstrates an important finding: unstructured assessment allows for an environment where executive dysfunction symptoms can consistently manifest.

Since the original MET was developed, several additional versions have been created, they include a hospital version, the MET-revised, the Baycrest MET, the revised Baycrest MET,

and the virtual MET. The hospital version of the MET was first used by Knight, Alderman and Burgess (2002) with traumatic brain injury and stroke patients. They modified the MET to include enhanced specificity about the rules. For example, the original MET had the rule "You are to spend as little money as possible (within reason)" (Shallice & Burgess, 1991, p. 734). The MET-Hospital Version corresponding rule is "You spend no more than $2.50" (Knight, Alderman, & Burgess, 2002, p.254). The MET-Hospital Version rule is therefore more straightforward and specific in regards to how much to spend. The MET-Hospital Version also has simplified task demands. For example, the original MET required participants to find information that could not be easily accessed such as "the name of the coldest place in Britain yesterday" (Shallice & Burgess, 1991, p. 733). Meanwhile, the information that is required to be found in the MET-Hospital Version is all related to the hospital setting where the test is administered, such as "What is the closing time of the staff library on a Friday?" (Knight, Alderman, & Burgess, 2002, p. 254). Lastly, the participants who completed the MET-Hospital Version were provided with an instruction sheet that directs participants to write down specific information. Overall, Knight and colleagues (2002) found that patients with brain injuries broke more rules and made more errors while successfully completing fewer tasks. Moreover, the changes that these authors made to the MET were beneficial in that both groups made very few errors of interpretation. This means that the findings described above can be confidently attributed to the differences in neurological status between the TBI group and the healthy group, not to errors in the participants' understanding the purpose of the assessment. Another benefit of this hospital version is that it demonstrated that the MET can be used with patients who cannot be assessed in public such as those with mobility issues, severe behavioral problems, or those detained due to mental health concerns.

Another hospital version of the MET was developed by Dawson and colleagues (2009) for use in a different hospital. The Baycrest version examined healthy participants as well as those with acquired brain injury due to stroke or TBI. The Baycrest MET tasks and rules closely resemble the MET-Hospital Version and another consistency is that an administration manual was created to assist administrators in performing the Baycrest MET and to ensure that all participants were given the same instructions. In addition, a standardized scoring method was created to increase clinical utility and interrater reliability. Using that scoring method, Dawson and colleagues (2009) found that those with acquired brain injury performed significantly worse than controls. Specifically, participants with acquired brain injury completed fewer tasks and demonstrated more overall errors than age and education-matched control participants. This demonstrated that the Baycrest MET was able to distinguish those who had suffered an acquired brain injury even after their completion of therapy and leaving of the hospital. Good to strong correlations (>.50) were found in over one third of the analyses examining the relationships between the Baycrest MET and measures of everyday function and instrumental activities of daily living (IADL). Good interrater reliability was also determined due to high interclass correlation coefficients on Baycrest MET summary scores (Dawson et al., 2009).

Soon after publishing their findings with the Baycrest MET, the research team at Baycrest revised their version to create the Baycrest Revised version. This new version was created to improve the original Baycrest MET's ability to discriminate between patients and healthy participants. Specifically, the Baycrest MET-Revised replaces an ambiguous Baycrest MET task with one that is more clear; while the Baycrest MET asks participants to "collect something for the examiner… and do what is necessary" (Dawson et al., 2009), the Baycrest MET-Revised introduces an interruption task 12 minutes into testing where participants are

instructed to find a specific flyer and hold onto it until they have completed the MET (Clark, Anderson, Nalder, Arshad & Dawson, 2017). The Baycrest MET-Revised also replaces a task that is not representative of real-life with one that is more representative and lastly, the "meet here after ten minutes of starting the exercise" was replaced with meeting somewhere at a specific time to clear up confusion if someone were to forget what time the assessment started. With this revision, Clark and colleagues (2017) found that patients with acquired brain injury (ABI) omitted more tasks, broke more rules, and were less efficient than healthy participants.

Despite the benefit of ecological validity, scoring of the original MET has been difficult for examiners. Recently, Morrison and colleagues (2013) addressed this shortcoming in the MET-Revised, a version in which both an observation form and simpler scoring system were integrated. The observation form allows for recording the number of locations visited while completing the MET-Revised tasks and the path that patients took to complete those tasks. To clear up the ambiguity of scoring, a simple scoring system is used with the MET-Revised where the ratio of tasks completed to total number of locations visited is expressed as performance efficiency. While this scoring method is certainly simpler, it also ignores other elements of participant performance such as not performing tasks as efficiently as possible, making mistakes due to a lack of understanding the instructions, breaking social rules.

The Virtual MET (VMET; Rand et al., 2009) is especially unique as virtual reality is utilized in order to administer the MET. This version was first used to assess patients after a stroke. The Virtual MET was modeled after the MET-Hospital Version but a few modifications were necessary to adapt to the technology (VMALL, a virtual reality shopping environment). The Virtual MET has the same number of tasks (items to be bought and information to be obtained) and identical instructions as the MET-Hospital Version. However, the items to be

bought were changed to those that could be created for the VMALL. Another minor modification involves the task where the patient is supposed to meet the assessor at a specific time. This task was changed to checking the contents of the shopping cart at a particular time and a digital clock is located at the bottom of the screen for this purpose (taking the place of watch in the MET-Hospital Version). To examine ecological and construct validity, participants in Rand and colleagues' (2009) study took the MET-Hospital Version and the Virtual MET. The Virtual MET was found to have moderate to high correlations with the MET-Hospital Version and was also able to discriminate between younger and older adults.

Upon initial discovery of the Virtual MET, one may think that it would not be as useful as the other MET versions due to its lack of ability to assess social errors, such as talking too long to people. However, with this version post-stroke patients had difficulty planning, multi-tasking, and problem solving, and they were also largely unaware of mistakes. The post-stroke patients made more total mistakes, mistakes in completing tasks, partial mistakes in completing tasks, non-efficiency mistakes, and use of strategy mistakes than both the older healthy adults and young healthy adults. Young healthy adults performed most efficiently, followed by older healthy adults, and then post-stroke patients. Therefore, the Virtual MET is able to successfully discriminate between those with and without executive dysfunction (Rand et al., 2009). The benefits to an assessment such as this include being safer, less time-consuming, and less expensive. The Virtual MET can be a good option particularly for those with motor difficulties as it minimizes the risk involved with possible injury while walking around doing the assessment. The Virtual MET is also potentially more convenient as the patient is not required to travel to take the assessment. This not only saves time for the patient, but also for the examiner, especially if the client is currently in a hospital/rehabilitation setting. This virtual version is also

beneficial because it can be used as a training tool. Similar to how the true-reality MET versions can be modified, the Virtual MET can be modified, such as changing the products to buy, so that patients can retake a slightly different version of it.

Ecological Validity

One feature the MET is highly praised for is its ecological validity, how it represents real-life activities and contexts (Burgess et al., 2006). This is important because it common for patients to complete the MET to predict what cognitive and behavioral challenges they will encounter outside of the hospital or rehabilitation center. For example, one who used to live alone before their injury may not be able to do so after frontal lobe damage because it is possible that they may no longer be able to safely prepare meals, engage in self-care, and/or adhere to a medicine regime. Moreover, even if the impairments are not that severe, a therapist can use MET performance to inform them about how to better prepare that patient for life. For example, a therapist may work with a patient about what strategies to use at the grocery store to be more efficient.

Maeir, Krauss, and Katz (2011) studied patients with acquired brain injury (e.g., traumatic brain injury, stroke, tumor), who were independent in basic activities of daily life one week before their discharge from a rehabilitation center. The patients in this study completed the MET at discharge and three months later, they completed the MET again as well as the Participation Index (M2PI) assessment, a separate assessment that determines community participation. Therapists had individuals and their caregivers complete the M2PI separately. The M2PI covers topics such as initiation, social contacts, leisure, self-care, residence, transportation, employment, money management, and provides a total participation index score. Many of these activities require executive function skills such as planning, socializing, and self-regulating. Maeir and colleagues (2011) found significant positive correlations between patients' MET

performance and their M2PI scores such that participants with low scores on the MET also scored low on the M2PI. Based off the M2PI, patients experienced a significant amount of issues with community integration after discharge, indicating that the MET is indeed an ecologically valid assessment. Because community participation is a major goal of occupational therapy, therapists can use MET performance to better prepare patients with strategies for being involved in the community.

The MET also demonstrates ecological validity in that it is correlated with other assessments that examine how one functions in everyday life. For example, it is moderately correlated with most of the self- and informant-reported items from the DEX, and also correlated with everyday functional ability as measured by the Assessment of Motor and Process Skills (AMPS) battery (Dawson et al., 2009). Finally, the ecological validity of the MET also makes it more seemingly relevant to patients (Morrison et al., 2013). It is common for patients to perform poorly on assessments that they feel are irrelevant. Patients are better able to clearly see their failures when they understand the relevance of the assessment.

Reliability

Dawson and colleagues (2009) demonstrated that the Baycrest MET version has high interrater reliability. High interrater reliability is important because this indicates that if two different therapists were to score a single participant's MET performance, they would make similar conclusions. Results need to be consistent, not dependent upon which therapist is giving the assessment. While Dawson and colleagues (2009) report high interclass correlation coefficients, the confidence intervals around the coefficients were large. In response, the authors suggest that a detailed instruction manual for scoring could be helpful in improving the interrater reliability (Dawson et al., 2009).

Clark and colleagues (2017) measured the test-retest reliability, administering the same assessment twice to the same person to ensure similar results, of their Baycrest MET-Revised by creating two slightly different versions. Tasks were similar; for example, purchase a drink in Version A and purchase a snack in Version B. The purpose of creating an alternate version of the Baycrest MET-Revised was for use as a potential outcome measure for further studies that examine the effectiveness of rehabilitation for executive dysfunction. Results indicated that both versions A and B of the Baycrest MET-Revised discriminated between acquired brain injury and healthy individuals.

Clark and colleagues (2017) also replicated high interrater reliability with the Baycrest MET-Revised. In fact, the Baycrest MET-Revised was found to have higher correlation coefficients than the original Baycrest MET. For example, the Baycrest MET had correlation coefficients of .80 for completions, .88 for total rule breaks, and .82 for total errors. The Baycrest MET-Revised had correlation coefficients of .98 and 1.00 for completions, .95 and .99 for total rule breaks, and .97 and 1.00 for total errors on Versions A and B, respectively. These higher reliability coefficients may be due to the modifications which decreased the ambiguity of a few of the tasks in the original Baycrest MET. The revisions eliminated tasks in which participants may have only made errors due to misunderstanding (Clark et al., 2017). Instead, the new tasks should more clearly be measuring errors due to executive dysfunction. With tasks that were less ambiguous, researchers could more clearly determine completions, rule breaks, and total errors of participants.

Applicability

A unique feature of the MET is that it can be modified according to the setting and population, making this an extremely flexible assessment. This allows therapists to test patients

in a way that is most beneficial to them. Due to the context-specific nature of the MET, it is not necessarily a one-size-fits-all assessment. It has been modified and revised multiple times for use in various settings and populations. In fact, the MET can be administered almost anywhere.

Therapists who use the MET report that it can be used in a variety of ways. Some therapists use the MET as a baseline assessment to determine the degree to which a patient's executive function is impaired. One occupational therapist described his/her use as, "I'll use it as just a pure assessment at the beginning to see where they're having trouble, particularly if they've done really well on other standardized assessments…" (Nalder et al., 2017, p. 695). It can also be used to determine a patient's readiness to go home from a rehabilitation center. This can assist in determining whether the patient can live alone or needs around-the-clock care, while also giving insight as to how the patient will function in the real world. The MET can also be used as a training tool for patients currently undergoing rehabilitation. As the patient completes tasks, the therapist can show them what they are doing wrong or inefficiently. The therapist can also lead patients toward helpful strategies for completing everyday activities, such as making a grocery list before going to the store and grouping the grocery items according to where they will be found. Some occupational therapists have patients take the MET, teach them strategies based on their performance, and then have them retake the MET to see if they are able to successfully apply those strategies (Nalder et al., 2017).

The MET can be utilized with a wide variety of patients. The most common examples include patients who experienced a traumatic brain injury and patients who have sustained a stroke. However, executive dysfunction can happen even without specific brain damage (Torralva, 2012). Indeed, the MET has also been used with those suffering from psychological disorders, such as bipolar disorder. In a study by Torralva et al. (2012), euthymic (normal, non-

depressed, reasonably positive mood) bipolar disorder patients who had typical scores on laboratory-based tests of executive function showed important deficits on the MET, specifically bipolar disorder patients broke more rules and were generally less efficient. This may have resulted from bipolar disorder participants acting impulsively, failing to plan ahead of time, and organizing poorly. This suggests that the MET can be used to detect executive impairments that are due to numerous conditions, even beyond those caused by brain injury.

The MET is also valued for its clinical utility, its ability to cause beneficial change in health outcomes (Nalder, Clark, Anderson & Dawson, 2017). Because of the MET's ability to predict life outside of the hospital, it is valuable to occupational therapists when creating treatment plans. Nalder and colleagues (2017) conducted interviews to determine the clinical utility of the MET from the perspective of clinicians. A popular, positive opinion of the MET was that it reflects real life, is complex, and is sensitive to high level cognitive deficits. Indeed, in discussing the merits of the MET, one occupational therapist stated, "I want complexity, messiness. I want multiple tasks, I want juggling of information and activities. I don't want structure, I don't want to be leading it, I don't want me to be guiding [patients] in any way. I want to know what they're like, completely on their own" (Nalder et al., 2017, pg. 692). Therapists can use patients' MET performance while in a treatment facility alongside information about a patient's environmental demands at home to determine how to safely return that patient to their home. If a patient only has a few executive function deficits but lives in a highly demanding environment, their everyday function may still be impacted and more support may be needed. However, if a patient has numerous executive function deficits but lives in an environment with fewer demands, their everyday function may not be greatly impacted and thus, less support may be needed (Chaytor, Schmitter-Edgecombe, & Burr, 2005).

Limitations

While the MET is certainly a beneficial tool for assessment, it does not come without its limitations. First, the MET can be costly because patients must often be transported to a different location and provided money to buy small items during the actual assessment. Second, the MET can be time consuming, particularly if the therapist plans on explaining errors, inefficiencies, and/or potential strategies for performance or if the assessment will take place outside of the hospital. This makes this assessment especially difficult for acute care patients who may only be in the hospital for a short amount of time. For these reasons, future research could include making a modified version of the MET for acute care, possibly a shortened version, or a technological tool that would use MET-like tasks to teach executive strategies. Third, the MET may not be safe for those with mobility issues. Completing an assessment in a public place where walking considerable distances is required would most likely be a risk that would not be worth taking for the sake of the safety of the patient as well as the liability on the therapist/hospital. Finally, clinicians have reported wanting more guidelines about how to best score MET performance. This scoring of the MET is the focus of the subsequent section.

Scoring the MET

When patients complete the MET, they are scored based on a number of characteristics. Inefficiencies are tabulated when a more effective strategy could have been applied. Rule breaks are counted when one of the nine specifically defined rules (within the test or social) are broken. Interpretation failures occur when the requirements of a task are misunderstood. Task failures are noted when any of the 12 tasks are not fully completed. Total failures is the total sum of errors

from all categories (Shallice & Burgess, 1991). Each of these variables is used to score the patient.

To generate these performance indicators, the MET is typically scored in two parts: 1) each task is scored based on whether it was completed accurately, partially, or not at all (omitted), and 2) other behaviors like rule breaks, strategy use, and inefficiencies are coded. As demonstrated in Table 1, the spreadsheet that is used to score the tasks features a place to write the order (by number) in which the tasks were completed, whether or not the task was completed (yes, no, or partially), the type of partial task failure if applicable, and whether that task failure was due to an inefficiency (failing to use time and resources in the best way possible) or due to an interpretation error (made an error due to misunderstanding of directions). Inefficiency errors and interpretation errors can be labeled as E and N, respectively, in order to save time and allow therapists to score as quickly as possible and not miss anything while writing.

Table 1

Order	Task	Y/PTF/N	If a PTF, select failure(s):	Inefficient/Interpret
1	Buy a get well card	PTF	__ Bought wrong kind of card __ Did not buy the least expensive card _x_ Spent excessive time selecting card (>2 min) __ Went back to card collection to exchange card	E

The behavior scoring spreadsheet lists the nine rules, and unlike the task scoring sheet where there can be partial task failures, the rules are either broken or not. However, some rules can be broken multiple times (such as going outside of the designated testing area boundaries)

and so the experimenter does keep track of how many times each rule is broken. In addition to the MET rules, the behavior scoring sheet also lists a variety of social rules that are commonly broken, a list of strategies that are often used, and a variety of behaviors that make MET performance more inefficient. Social rule breaks include behaving abnormally like being rude, interrupting someone while they are speaking, or getting in front of someone already in line. It is important to note that for the Virtual MET, there is not a social rule break section as one is not actually able to interact socially through virtual reality.

The strategies category examines the actions the participant takes in order to successfully complete the MET. The experimenter observes if the participant uses a map, asks a staff member for help, and/or plans before starting the test. If a patient does not voluntarily use these types of strategies, that information can be used as a teaching tool for the patient's therapist by teaching potential strategies to improve and even testing patients at a later date to determine if they are effectively use the new strategies.

As the name suggests, inefficiencies are noted for behaviors that participants demonstrate are not related to successful completion of the MET and that keep the participant from being as efficient as possible, in regards to both time, strategy, and money. An example of an inefficiency is a participant seeing a friend and having a ten-minute conversation with them. While it would be abnormal for the participant to simply ignore or avoid their friend, this is an inefficient behavior due to the excessive amount of time spent conversing. Other examples would include asking the same question multiple times, sitting down without a good reason, and asking random people for help (instead of staff). Other MET researchers have also tracked behaviors like the number of times the participant checks the task list while walking versus while stopped, the number of times the participant checks his/her watch, and the number of times the participant

engages in self-talk. This can be quite onerous to keep up with, so it is important for the therapist or experimenter to determine whether or not keeping track of the detailed information will be useful for him/her in the therapy process or in answering the research question.

The scoring of the MET can be a daunting process if one is not properly prepared. Therefore, it can be discouraging to some therapists wishing to utilize the assessment. Many therapists are more comfortable with structured, normative assessments, those with clear classifications of what is typical (Nalder et al., 2017). Indeed, therapists may hesitate to use the MET because it lacks structure and a set of normative data. This is paradoxical because what makes the MET useful for assessing executive function is its lack of structure. Therapists wanting to use the MET will need to be flexible but improvements can be made to make scoring the MET more clear and consistent for all those who use it.

A common challenge therapists have faced when scoring the MET has been difficulty in characterizing normal performance (Nalder et al., 2017). In fact, Nalder and colleagues' (2017) reported that, "all clinicians felt there needed to be better characterization of 'normal performance' to assist in interpreting the MET" (p.698). One example of a potentially confusing error is found in the task of purchasing a birthday card. If the participant purchases a wedding card instead, is this a mistake anyone could accidentally make or is this error cause for intervention? These are the types of questions therapists routinely encounter when scoring the MET. For this reason, many therapists feel more confident in using the MET as a teaching tool (strategies to use in everyday life situations) instead of a predictor of abilities outside of the hospital (Nalder et al., 2017). However, there are solutions to helping solve this issue. This includes proper training of the therapists giving the assessment through organized workshops where attendees watch videos of participants completing the MET, score the videos, and discuss

them with experts. In addition, future research will focus on collecting normative data so that therapists can better interpret which patterns of behavior fall outside of normal limits. This will help determine if an error is indicative of executive dysfunction or perhaps it is an error that is commonly made by healthy participants as well. Finally, when scoring and interpreting MET performance it is important for therapists to keep in mind that the overall performance of the assessment is more indicative than any single error or behavior. Therapists should look for consistency of errors and disorganization in performance instead of focusing too much attention on any one element of the task.

The fact that the MET is scored with both quantitative and qualitative measures is also beneficial. Quantitative data, such as number of rule breaks, is concrete data that can be recorded about a patient. It is useful data because it can serve as proof of improvement. A patient can be tested with the MET, and the scores can be recorded. The therapist can then use treatment sessions to demonstrate how to improve and types of strategies to implement. Several months later, the therapist may have the patient complete the MET again. The new scores can be compared to the old scores, and this is an easier way to show improvement versus simply stating that it seemed to go better. However, the qualitative information serves an important purpose as well. This is descriptive information of how the patient carried out the tasks, specifics of organization and approach. In fact, many researchers find this information quite useful for creating and evaluating treatment plans (Nalder et al., 2017). Clinicians can note areas where patients were not efficient and later use that information to create a plan for progress.

Despite these benefits, the scoring procedures associated with most versions of the MET can be cumbersome, the MET-Revised was created with a simplified scoring technique in mind. Morrison and colleagues (2013) created a short observation form to simply record the number of

locations a participant visited as well as the path that participants took during the task. With the only four dependent variables are collected: total time to complete the assessment, number of locations visited, number of tasks completed, and number of rule breaks (sum of all types of rule breaks and their frequencies). With these data points, the experimenter/therapist calculates the participants' performance efficiency, the ratio of tasks completed to total number of locations visited while completing the MET-Revised. This takes into account how many places the participant must visit in order to accomplish all of the tasks. For example, participant 1 and participant 2 both bought the required card and pack of gum. However, participant 1 bought both of the items at the gift shop, while participant 2 bought the card at the gift shop but the gum at the vending machine. Participant 1 would have a higher performance efficiency in this example.

The revised scoring approach used by Morrison and colleagues (2013) was associated with very high interrater reliability as the two raters in this study had identical scores. While this simplified scoring approach may be easier, it does come with drawbacks. The unique attributes of the MET come from its unstructured and unpredictable setting and its flexibility to allow each participant/patient to demonstrate their real-world skills. If therapists and experimenters all but ignore use of strategies, examples of inefficiencies, and other peculiar behaviors, a lot of valuable information is missed. Consequently, while the MET-Revised approach may ease the scoring process, it also views executive dysfunction in a narrower manner. Because of its benefits as well as drawbacks, therapists and experimenters should take careful consideration of which MET scoring approach is best for his/her patient or research goals.

Reflection

Another element of my thesis involved working with Life Care of Hixson to assist them in making their own version of the Multiple Errands Test (Life Care Centers of America-MET).

Life Care of Hixson had already started this process before I contacted them. I was able to help them modify the instructions and tasks they had already created (decreasing the number of items to be bought, adding a multi-tasking component, editing questions to be more related to executive function, etc.). In addition to assisting with the instructional and task component, I created a behavior and task scoring sheet specific to the Life Care Centers of America-MET. These scoring sheets listed mistakes that could be made, specific to this version. This process took some trial and error. I created the scoring sheets, met with Josh Haislip (Director of Rehabilitation), Matthew Dudek (Assistant Director of Rehabilitation), and Rachel Halter (Occupational Therapist) of Life Care of Hixson, and then made changes to create the final drafts of the scoring sheets. In late March, I will test the Life Care Centers of America-MET, most likely with a healthy, college student to ensure that the test can be clearly understood. While I will likely be graduated by the time Life Care of Hixson can actually use this assessment with patients, I am hopeful that another UTC student will continue assisting with the progression of the Life Care Centers of America-MET. The scoring sheets I created have been included at the end of this chapter, as well as the instructions that I assisted in modifying. It has been a rewarding experience to be able to actually apply the MET knowledge I have gained to making a site-specific version.

Future Directions

While the MET has been a useful tool for measuring executive dysfunction, there are still many areas of research to be explored with respect to this assessment. Future research may examine the MET's relevance in the aging process. Tests of executive function have predicted instrumental activities of daily living in older individuals (Cahn-Weiner et al., 2002). One study, while the main purpose was testing those with mild cerebrovascular accident, used the MET to

discriminate between the younger and older controls. Findings such as this could encourage future research on using the MET as a tool to measure declines due to aging, such as determining appropriate scores for participants based on age and the type of executive function changes expected based on age. This could help determine aging adults who may be declining and need additional assistance.

Another future direction for the MET is to measure the associations between rule breaks on the MET and real-life behaviors. For example, in Alderman et al.'s study (2003), patients who broke more rules showed increased problems with control aspects of memory (confabulation, temporal sequencing problems, repetitive behavior). Those who failed to complete the set tasks were linked to negative symptoms such as apathy and shallow effect (shallow, short-lived emotions) (Alderman et al., 2003). Information such as this could be useful for occupational therapists when creating treatment plans. This knowledge could also be beneficial to the potential caretakers of the patient, such as knowing you may need to provide assistance with tasks involving memory. If a patient is showing apathy, this could also be helpful information for caretakers to be aware of. Future research could expand upon this research in finding these associations.

Summary

Learning about the MET has been a highly rewarding experience for me. After some discussion with my thesis director, Dr. Clark, I decided on this topic due to my interest in occupational therapy and psychology, my undergraduate major. In Fall 2018 will be attending graduate school for my Occupational Therapy Doctorate (OTD) degree. I wanted to research and write about something that could possibly be useful to me in the future, as well as something that is interesting to me. In the future, I will most likely be working with patients with executive

dysfunction whether they be patients with TBI, some type of addiction, or even a child with ADHD. I am thankful for my newfound knowledge of executive function and the MET as I look forward to applying this knowledge within my profession.

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. doi:10.1017/s1355617703910046

Army Individual Test Battery. (1944). *Manual of directions and scoring*. Washington, DC: War Department, Adjutant General's Office.

Benton, A. L., Hamsher, K., & Sivan, A. B. (1983). *Multilingual aphasia examination* (3rd ed.). Iowa City, IA: AJA Associates.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., M-A. Coates, L., 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. doi:10.1017/s1355617706060310

Cahn-Weiner, D. A., Boyle, P. A., & Malloy, P. F. (2002). Tests of executive function predict instrumental activities of daily living in community-dwelling older individuals. *Applied Neuropsychology, 9*(3), 187-191. doi:10.1207/s15324826an0903_8

Cahn-Weiner, D. A., Malloy, P. F., Boyle, P. A., Marran, M., & Salloway, S. (2000). Prediction of functional status from neuropsychological tests in community-dwelling elderly individuals. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D), 14*(2), 187-195. doi:10.1076/1385-4046(200005)14:2;1-z;ft187

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology, 21*, 217-227. doi:10.1016/j.acn.2005.12.002

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. doi:10.1080/09602011.2015.1117981

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,* S41-S51. doi:10.1016/j.apmr.2009.07.012

Dywan, J., & Segalowitz, S. (1996). Self and family ratings of adaptive behavior after traumatic brain injury: Psychometric scores and frontally generated ERPs. *Journal of Head Trauma Rehabilitation, 11*(2), 79–95.

Eslinger, P. J., & Damasio, A. R. (1986). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology, 35*, 1731-1741. doi:10.1097/00001199-198609000-00012

Golden, C. (1978). *Stroop color and word test: A manual for clinical and experimental uses*. Los Angeles, CA: Western Psychological Services

Eling, P., Derckx, K., & Maes, R. (2008). On the historical and conceptual background of the Wisconsin Card Sorting Test. *Brain and Cognition, 67*, 247-253. doi:10.1016/j.bandc.2008.01.006

Gillen, Glen. *Cognitive and Perceptual Rehabilitation: Optimizing Function*. Mosby, Inc., 2009.

Goldstein, S., & Naglieri, J. A. (2014). *Handbook of Executive Functioning.* New York, NY: Springer. doi:10.1007/978-1-4614-8106-5

Grant, D. A., & Berg, E. A. (1948). A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology, 38,* 404–411.

Jodzio, K., & Biechowska, D. (2010). Wisconsin Card Sorting Test as a measure of executive function impairments in stroke patients. *Applied Neuropsychology, 17*(4), 267-277. doi:10.1080/09084282.2010.525104

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. doi:10.1080/09602010244000039

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). doi:10.3928/15394492-20101108-07

Malec, J. F. (2004). Comparability of Mayo-Portland Adaptability Inventory ratings by staff, significant others and people with acquired brain injury. *Brain Injury, 18,* 563-575.

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. *The American Journal of Occupational Therapy, 67*(4), 460-468. doi:10.5014/ajot.2013.007880

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. doi:10.1080/09602011.2015.1067628

Penfield, W. & Evans, J. (1935). The frontal lobe in man: A clinical study of maximum removals. *Brain, 58,* 115–133.

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

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B, 298,* 199–209.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727-741. doi:10.1093/brain/114.2.727

Torralva, T., Strejilevich, S., Gleichgerrcht, E., Roca, M., Martino, D., Cetkovich, M., & Manes, F. (2012). Deficits in tasks of executive functioning that mimic real-life scenarios in bipolar disorder. *Bipolar Disorders, 14,* 118-125. doi:10.1111/j.1399-5618.2012.00987.x

Wilson, B. A., Alderman, N., Burgess, P., Emslie, H., & Evans, J. J. (1996). *Behavioural assessment of the dysexecutive syndrome (BADS)*. Bury St. Edmunds, England: Thames Valley Test Company.

Appendix A

LCCA-MET – TASK SHEET

ORDER	TASKS	(Y/PTF/N)	If a partial task failure (PTF), select failure(s):	inefficient/interpret
	Buy 1 stamp		___ Bought more than 1 stamp ___ Bought something other than the required items ___ Spent excessive time selecting items (>2 min.)	
	Buy 1 Sprite		___ Bought more than 1 Sprite ___ Bought a beverage other than Sprite ___ Bought something other than the required items ___ Spent excessive time selecting items (>2 min.)	
	Buy a get well card		___ Bought wrong kind of card ___ Did not buy the least expensive card ___ Spent excessive time selecting card (>2 min.) ___ Went back to card collection to exchange card	
	Call the Director of Rehab (423) 591-6650 and leave your name, what time it is, and the day of the week. CHECK IF CORRECT INFORMATION GIVEN_____		___ Changes telephones unnecessarily ___ Needs to hang up and try again ___ Called wrong person ___ Gave incorrect information in telephone call ___ Gave too little information in the telephone call ___ Gave too much information in telephone call	
	Mail Get Well Card to your address ADDRESS WRITTEN CORRECTLY_____		___ Forgot to put stamp on card ___ Wrote address incorrectly ___ Left item with examiner/other individual ___ Gave to examiner AFTER test was finished	
	What activity happens on Saturday at 12 PM? CHECK IF ACTIVITY IS CORRECT_____		___ Wrote incorrect activity ___ Wrote correct activity only by chance	
	What is on the menu for Lunch today? CHECK IF DINNER ITEM IS CORRECT_____		___ Wrote incorrect menu item ___ Wrote correct menu item only by chance	
	Who is the Executive Director for the facility? CHECK IF CORRECT_____		___ Wrote incorrect name ___ Misspelled name	
	Tell the therapist that you have completed the whole task		___ Did not state completion ___ Told someone other than the therapist	
	Interruption Task		___ Did not introduce themselves ___ Forgot to record the name of the nurse	

TOTALS: # YES:_____/10 # Partial Task Failures:_____ #NO:_____/10

Appendix B

LCCA BEHAVIOR SCORING SHEET

RULE BREAKS

__ Failed to complete all the tasks
__ Spent more money than necessary
__ Went into a treatment or "staff only" area
__ Went outside the main floor of the hospital
__ Returned to an area
__ Bought more than 2 items at one place
__ Wasn't as efficient as possible
__ Rushed excessively
__ Spoke to the experimenter unnecessarily

TASK UNRELATED

__ Wandered around an area
__ Unexplained hesitations during task (>20sec)
__ Unnecessary conversations with individuals >20sec)
__ Entered / Attempted to enter an unnecessary area (without any clear reason)
__ Visited an area & didn't do anything there
__ Unnecessarily sat down
__ Repeated question(s)
__ Took obvious steps to looking for unnecessary information/items
__ Asked non-staff for directions / help / assistance

SOCIAL RULE BREAKS

__ Remained at checkout unnecessarily
__ Left item(s) behind
__ Interacted rudely (verbally or non-verbally)
__ Interrupted someone
__ Went in front of someone in line
__ Entered an area via a clearly marked exit

STRATEGIES

__ Planned before starting test
__ Used a map
__ Made notes / Marked tasks as completed
__ Self-Talk (Task Oriented)
__ Self-Talk (Non-Task Oriented)
__ Asked Staff for directions / help / assistance
__ Looked overtly at signage / visual landmarks
__ Went to meeting place early & waited
__ Engaged in multitasking
__ Organized materials and bag
__ Checked watch