Ecological validity of the Multiple Errands Test using predictive models of dysexecutive problems in everyday life

Gustavo Cuberos-Urbano, Alfonso Caracuel, Raquel Vilar-López, Carlos Valls-Serrano, Andrew Bateman, and Antonio Verdejo-García

Department of Clinical Psychology, Universidad de Granada, Granada, Spain
Department of Developmental Psychology, Universidad de Granada, Granada, Spain
Institute of Neuroscience F. Olóriz, Universidad de Granada, Granada, Spain
Oliver Zangwill Centre for Neuropsychological Rehabilitation, The Princess of Wales Hospital, Ely, UK
NIHR CLAHRC for Cambridgeshire and Peterborough, Cambridge, UK

Published online: 04 Mar 2013.


To link to this article: http://dx.doi.org/10.1080/13803395.2013.776011

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Ecological validity of the Multiple Errands Test using predictive models of dysexecutive problems in everyday life

Gustavo Cuberos-Urbano1, Alfonso Caracuel2,3, Raquel Vilar-López1,3, Carlos Valls-Serrano1, Andrew Bateman4,5, and Antonio Verdejo-García1,3

1Department of Clinical Psychology, Universidad de Granada, Granada, Spain
2Department of Developmental Psychology, Universidad de Granada, Granada, Spain
3Institute of Neuroscience F. Olóriz, Universidad de Granada, Granada, Spain
4Oliver Zangwill Centre for Neuropsychological Rehabilitation, The Princess of Wales Hospital, Ely, UK
5NIHR CLAHRC for Cambridgeshire and Peterborough, Cambridge, UK

The “dysexecutive syndrome” is composed of a range of cognitive, emotional, and behavioral deficits that are difficult to evaluate using traditional neuropsychological tests. The Multiple Errands Test (MET) was originally developed to systematize the assessment of the more elusive manifestations of the dysexecutive syndrome. The aims of this study were to examine the reliability of the MET and to investigate the predictive ability of its indices to explain a range of “dysexecutive”-related symptoms in everyday life. Thirty patients with acquired brain injury participated in this study. The MET showed an adequate inter-rater reliability and ecological validity. The main performance indices from the MET were able to significantly predict severity of everyday life executive problems, with different indices predicting particular manifestations of different components of executive functions.

Keywords: Neuropsychology; Dysexecutive syndrome; Ecological validity.

INTRODUCTION

Executive functions are a key asset for the achievement of autonomy and independence in activities of daily living, and their impairment is a major determinant of poor functional outcomes after acquired brain injury (ABI; Reid-Arndt, Nehl, & Hinkebein, 2007). Damage to the frontal lobes and frontal-striatal systems may lead to a diverse set of symptoms often referred to as the dysexecutive syndrome (Hanna-Pladdy, 2007). This dysexecutive syndrome is actually composed of an array of clinical manifestations that include cognitive (e.g., inattentiveness, distractibility, or poor planning), emotional (e.g., apathy or lability), and behavioral symptoms (e.g., stimulus-bound behavior, disinhibition, perseveration; Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Cummings, 1993). The complexity of these manifestations often means that traditional paper-and-pencil neuropsychological tests are insufficient to cover all the relevant domains and dimensions of the dysexecutive syndrome (Fortin, Godbout, & Braun, 2002). In addition, while traditional neuropsychological tests are focused on the detection of cognitive deficits, it is known that alterations in behavioral control and emotion regulation can readily lead to deterioration in everyday functioning even in the absence of cognitive impairment (Alexander & Stuss, 2006). As a result of these limitations, some patients...
with ABI may perform relatively well on standard neuropsychological assessments, but still display a range of dysfunctional behaviors when observed in unstructured everyday life situations characterized by multitasking demands, multiple sources of interference, and constraints imposed by social rules (Levine, Dawson, Boutet, Schwartz, & Stuss, 2000).

In an attempt to systematize the assessment of the “dysexecutive syndrome,” to measure some of their critical components that are not detected by traditional neuropsychological tests, and to increase the ecological validity of executive functions assessment, Shallice and Burgess (1991) developed the Multiple Errands Test (MET) as a multitasking probe that requires performing multiple fairly simple but open-ended tasks in a shopping context. This test has been shown to be more sensitive than traditional neuropsychological measures to elicit and detect executive failures such as distractibility, inefficient task implementation, or break-up of task and social rules. It has also been shown to be better at predicting behavioral problems in everyday life (Alderman, Burgess, Knight, & Henman, 2003). On the other hand, a classical critique to this type of test is the risk of losing internal validity, due to reduced psychometric strength. This critique has been also counterargued by evidence of strong inter-rater reliability (Dawson et al., 2009; Knight, Alderman, & Burgess, 2002), but this issue should be still taken into account when evaluating the pros and cons of applying this ecological approach to assessment and prediction of clinical outcome in clinical neuropsychology studies.

None of the available studies has examined the capacity of specific MET performance indices to predict behavioral manifestations of damage to different frontal-striatal systems. The Frontal Systems Behavioral Scale (FrSBe; Grace & Malloy, 2001) is a well-validated measure to assess the clinical manifestations of damage to three dissociable neural functional circuits that may provoke “dysexecutive syndrome” related symptoms: apathy—related to the medial frontal-thalamic circuit; disinhibition—related to the orbitofrontal cortex-thalamic circuit; and dysexecutive—related to the dorsolateral prefrontal cortex-thalamic circuit. We reasoned that different MET performance indices may differentially contribute to predict separate aspects of the “dysexecutive syndrome” manifestations; specifically, task inefficiencies should relate to apathy and executive problems, whereas rule break-up should relate to disinhibition problems. Therefore, the aim of this study was twofold: (a) to test the reliability of the MET, and (b) to examine the ability of MET-derived indices to predict different “dysexecutive”-related symptoms of apathy, disinhibition, and disorganized behavior in everyday life, as reported by relatives of those with ABI.

**METHOD**

**Participants**

Thirty individuals (25 men and 5 women) with acquired brain injury (19 due to traumatic brain injury, 8 to stroke, 2 to tumor, and 1 to anoxia) volunteered to participate in this study. Participants were recruited from the Department of Rehabilitation at the “Hospital de Rehabilitacion y Traumatologia” in Granada, Spain (HRT-Granada), and they were community dwelling, living alone or with their families. Patients were eligible if they were at least 5 months post injury, were able to read, understand, and speak Spanish, and were at least 18 years old. Participants underwent a broad neuropsychological assessment to rule out severe cognitive deficits. Magnetic resonance imaging scan was used to determine the damage location. All of the patients showed frontal damage (35.5% right damage, 35.5% left damage, and 29% bilateral damage); also, 77.4% had a temporal injury, 51.6% parietal, and 16.1% occipital, and 25.8% of the patients suffered diffuse axonal injury. Age range was 19–60 years (mean = 34.76; SD = 13.4), education range was 6–19 years (mean = 10.93; SD = 3.89), and mean time since injury was 63.6 months (SD = 40.2; range 5–153). The ethical committee of the hospital approved the study and authorized the video recording of the patients performing the test, and the staff implicated agreed to collaborate. All of the participants provided informed consent after receiving the information about the objectives and characteristics of the study.

**Instruments**

The Multiple Errand Test Hospital Version (MET-HV; Knight et al., 2002) was translated to Spanish, maintaining the structure and basic characteristics of the original. Similar to the original version, participants were required to achieve four sets of simple tasks, comprising 12 subtasks. Nevertheless, we included some variations to adapt the test to the cultural, economic, and physical environment of the HRT-Granada hospital (i.e., buy any cola soft drink instead of a Coca-Cola). Due to the environmental characteristics of the HRT hospital (e.g., there were more places where the same task could be performed, locations were spread across
five different buildings, and the participant had to visit several outdoor spaces, etc.), the Spanish version of the MET is more similar to that of the St. Andrew’s Hospital version (Knight et al., 2002) than to that of the Baycrest Centre for Geriatric Care (Dawson et al., 2009). After conducting the necessary adjustments for the Spanish version (see Appendices A and B), a pilot study with three healthy volunteers was conducted to ensure that changes on the test demands were well adjusted to the testing context and to make sure that no test-unrelated difficulties arose during test performance in the hospital. As defined by Shallice and Burgess (1991), the selected dependent variables related to error types were (a) inefficiencies (a more effective strategy could have been applied), (b) interpretation failures (requirements of a particular task are misunderstood), (c) task failure (a task is either not carried out or not completed), (d) actual rule breaks, (e) social rule breaks, and (f) total errors, which were weighted according to the weights originally used by Alderman et al. (2003). Weighted error scores demonstrated more sensitivity to neurological damage than direct error scores, applying these transformations: Errors that were seen in up to 95% of healthy controls were given a score of 1; errors only demonstrated by 5% or less of controls were assigned a score of 2; and errors unique to patients were given a weighted score of 3 (Alderman et al., 2003). All of the errors that patients committed in the present study were included in the examples published by Alderman et al. (2003). This finding points to the suitability of that list of errors to be used with different pathologies, environments, and cultures, and it supports the Alderman weighted scoring system. In addition, strategies (frequency with which they looked at the map, read signs, requested help, and performed multiple tasks, and the time they took to plan before starting) were registered according to Knight et al. (2002).

The Frontal System Behavioral Scale (FrSBe; Grace & Malloy, 2001), Spanish version (Caracuel et al., 2008), was completed by relatives who were in close contact with the patients’ everyday activities. The FrSBe contains 46 items that assess behavioral problems linked to dysfunction on three separated frontal-striatal systems: medial frontal (apathy), orbitofrontal (disinhibition), and dorsolateral prefrontal (executive dysfunction). The instrument yields scores from three independent subscales, which were used as dependent variables.

The Dysexecutive Questionnaire (DEX; Wilson, Alderman, Burgess, Emslie, & Evans, 1996) was also completed by the patients’ relatives. This questionnaire contains 20 items that assess a number of cognitive, emotional, and behavioral deficits typically found in patients with dysexecutive symptoms, including intentionality/planning, perseveration, or decision-making problems. The five factor-scores obtained by Burgess et al. (1998) were used as dependent variables (inhibition, intentionality, executive memory, and positive and negative affect).

Procedure

The test began at the entrance of the hospital, where participants received the material to perform the tasks: a pen, a 10-euro note, a carrier bag, a watch (if participants did not have one), a clipboard with the test instructions, and a map of the hospital including legends and details of the different buildings and possible locations to fulfill the tasks (see Appendices A and B). Instructions were carefully debriefed to the participants, who were also advised to check out the written instructions provided in the clipboard. At the end of debriefing, participants were prompted to ask questions about the test instructions, and, once resolved, they were requested to provide their own explanation of the tasks they had to do and the rules to complete them. Instructions were repeated if needed until we had made sure that participants perfectly understood the demands and the rules of the test. Self-efficacy (self-perception of the ability to carry out the tasks of the test) and familiarity with the hospital environment were registered before the beginning of the test, and a valuation of the patients’ rating own performance on the task was registered after the test was finished.

Each participant was followed at a certain distance by the examiner, who video-recorded patients’ performance for subsequent video examination and scoring. Two independent assessors with different levels of expertise (the examiner was a licensed psychologist, and the other assessor was a PhD neuropsychologist) rated each patient’s performance on the test by examining the video recordings. These ratings served us to calculate the inter-rater reliability of the test. After obtaining sound inter-rater reliability data, the examiner scores were used for all the analysis.

Statistical analyses

Descriptive statistics were computed for the measures of the MET-HV. Intraclass correlation coefficients (ICCs) were obtained to determine the inter-rater reliability. Finally, to examine the ecological validity of the MET-HV, and its ability to predict different aspects of the “dysexecutive syndrome,”
we conducted multiple linear regression models including the MET-HV performance indices as predictor variables and the relatives-informed behavioral problems (measured by the FrSBe and the DEX) as dependent variables. Analyses were conducted on SPSS Version 17.

RESULTS

Performance on the HRT-Granada adaptation of the MET-HV, the DEX, and the FrSBe

A summary of descriptive statistics for the sample on the different measures of the MET-HV, including error and strategies scores, are displayed in Table 1. As shown in the table, actual rule breaks and task failures were the more frequently committed errors, while requesting help and looking at the map were the more frequently used strategies.

Table 1 also displays descriptive scores for the different subscales of the FrSBe and DEX inventories.

No significant correlations were found between the MET scores and the self-efficacy, familiarity with the hospital environment, or the valuation of the patients’ rating of their own performance on the task.

Inter-rater reliability

ICCs were derived from two-way analyses of variance, using two scores per participant (one per assessor). ICCs were evaluated on four measures pertaining to errors (total tasks completed accurately, actual rule breaks, social rule breaks, and total errors—a composite measure of inefficiencies + misinterpretations + total task failures + actual rule breaks + social rule breaks) and two pertaining to the most frequent strategies (total multiple tasks, total requests for help). The obtained inter-rater reliability coefficients were adequate, ranging from .85 to .98 (see Table 2).

Ecological validity of the HRT-Granada MET-HV

The multiple regression models included the weighted scores of the HRT-Granada MET-HV performance indices (rule breaks and task failures) as predictor variables and the FrSBe and DEX

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Descriptive statistics for the sample on the measures of the MET, and FrSBe and DEX scores rated by the patients’ relatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>Variable</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>MET-HV errors</td>
<td>Actual rule breaks</td>
</tr>
<tr>
<td></td>
<td>Task failures</td>
</tr>
<tr>
<td></td>
<td>Social rule breaks</td>
</tr>
<tr>
<td></td>
<td>Inefficiencies</td>
</tr>
<tr>
<td></td>
<td>Interpretation failures</td>
</tr>
<tr>
<td></td>
<td>Total errors</td>
</tr>
<tr>
<td>MET-HV strategies</td>
<td>Frequency looked at map</td>
</tr>
<tr>
<td></td>
<td>Frequency read signs</td>
</tr>
<tr>
<td></td>
<td>Frequency multiple tasks</td>
</tr>
<tr>
<td></td>
<td>Frequency requested help</td>
</tr>
<tr>
<td></td>
<td>Initial planning time (seconds)</td>
</tr>
<tr>
<td>FrSBe</td>
<td>Apathy</td>
</tr>
<tr>
<td></td>
<td>Disinhibition</td>
</tr>
<tr>
<td></td>
<td>Dysexecutive</td>
</tr>
<tr>
<td>DEX</td>
<td>Inhibition</td>
</tr>
<tr>
<td></td>
<td>Intentionality</td>
</tr>
<tr>
<td></td>
<td>Executive memory</td>
</tr>
<tr>
<td></td>
<td>Positive affect</td>
</tr>
<tr>
<td></td>
<td>Negative affect</td>
</tr>
</tbody>
</table>

Note. MET-HV = Multiple Errands Test–Hospital Version; FrSBe = Frontal System Behavior Scale; DEX = Dysexecutive Questionnaire; SD = Standard deviation; Min = minimum; Max = maximum.
scores as dependent variables. Only two predictors were selected in order to maintain the balance between the number of predictors and the sample size (N = 30, at least 10 cases by predictor; Hair, Black, Babin, Anderson, & Tatham, 2006). We selected the “rule breaks” and “task failures” indices because they were the most frequently committed errors in this sample and because they have shown specific association with executive problems in everyday life in previous studies (Alderman et al., 2003; Knight et al., 2002). Both variables showed normal distribution and absence of collinearity.

The regression models to predict the FrSBe subscales scores were statistically significant for apathy (percentage of explained variance = 24%, p = .010, best predictor “task failures”), disinhibition (percentage of explained variance = 31%, p = .002, predictors “rule breaks” and “task failures”), and executive dysfunction (percentage of explained variance = 23%, p = .011, best predictor “task failures”; Table 3). In summary, HRT-Granada MET-HV indices can predict FrSBe scores, with percentages of explained variance ranging from 23% (executive dysfunction) to 31% (disinhibition). Overall, the best predictor is the index of “task failures,” although the index of “rule breaks” is also a significant predictor of disinhibition problems.

The regression models to predict the DEX scores were statistically significant for intentionality (percentage of explained variance = 24%, p = .009, predictors “task failures” and “rule breaks”). Therefore, the HRT-Granada MET-HV indices can significantly predict only the DEX measure of intentionality (see Table 3).

Regression models were replicated using direct scores (instead of weighted scores), and results remained stable.

### DISCUSSION

The aims of the present study were to explore the inter-rater reliability and ecological validity of the MET-HV by examining a population of brain-lesioned patients and a novel outcome measure of dysexecutive problems. Our results showed that, despite adequate understanding of the test demands, the MET-HV was able to elicit and reliably measure a number of errors associated with dysexecutive problems, including task inefficiencies, rule breaks, and inadequate use of available strategies. The most frequent errors were “rule breaks,” followed by “task failures,” a finding that is in agreement with results of previous studies conducted in the UK and the US (Alderman et al., 2003; Knight et al., 2002). Furthermore, by using the proposed cutoff of 7 or more total errors (Knight et al., 2002), 87% of our sample would be correctly classified as patients with “dysexecutive” problems, a finding that highlights the sensitivity of the test to detect the elusive manifestations of the “dysexecutive syndrome,” which are often neglected by traditional tests.

Taking into account that the MET-HV is considerably less structured than typical clinical neuropsychological tests, the examination of inter-rater reliability becomes especially relevant (Knight et al., 2002). Our results showed ICCs above .85,
demonstrating adequate inter-rater reliability for this adaptation of the test, in agreement with data from previous versions of the MET (Dawson et al., 2009; Knight et al., 2002).

In addition, our results further support the ecological validity of the HRT-Granada MET-HV by demonstrating that patients’ performance on the test can predict the behavioral problems that they experience in everyday life, as rated by close relatives in the FrSBe and DEX scales. One of our main aims in this regard was to examine whether different MET-HV indices could predict particular manifestations of frontal-striatal systems damage, measured by the FrSBe scale. Results showed that both apathy and executive dysfunction manifestations were predicted by “task failures,” whereas disinhibition symptoms were additionally predicted by “rule breaks.” This finding is in agreement with the assumption that impaired inhibition is intimately connected with the inability to self-regulate behavior in accordance with complex rules (Levine et al., 2000; Stuss & Alexander, 2007). The robust capacity of the HRT-Granada MET-HV indices to predict disinhibition problems (31% of explained variance) is especially relevant, given the difficulty in assessing cognitive and emotion dysregulation symptoms employing traditional neuropsychological tests (Nigg, 2000) and the relevance of these deficits to predict social adjustment (Malloy, Bihrlle, Duffy, & Cimino, 1993). On the other hand, the “task failures” index seems to work as an overall predictor of a broad spectrum of problems related to the “dysexecutive syndrome,” including some aspects that can be reliably measured by classic psychometric tools (e.g., poor planning or inflexibility) and others that are more elusive to traditional measures (e.g., apathy).

For the DEX measures, intentionality was related to both “task failures” and “rule breaks,” indices that showed associations with DEX scores in previous studies (Alderman et al., 2003; Dawson et al., 2009; Knight et al., 2002). Again, relevance of this data relies on the capacity of the MET-HV to substantially predict a dimension of executive dysfunction related to planning or goal-directed behavior. Research with different instruments and cultures has demonstrated that intentionality is a central component of executive functions (Amieva, Phillips, & Della Sala, 2003; Chan, 2001; Chaytor & Schmitter-Edgecombe, 2007; Simpson & Schmitter-Edgecombe, 2002). The fact that the MET predicts only intentionality supports the importance of this component for executive functioning in everyday life behaviors. The DEX items included in this factor refer to planning difficulties, lack of insight, knowing-doing dissociation, distractibility, and poor decision making. Thus, these executive subprocesses seem to be some of the main demands that the MET poses to patients. From a theoretical perspective, the findings of the present study held the existence of “intentionality” as a component of the executive functions, which is not easily detected using traditional measures but which plays a critical role in the clinical management of brain-damaged patients (Stuss & Alexander, 2007; Stuss & Levine, 2002).

Our study presents some limitations, the most important being the small sample size and the absence of a control group. Also, one weakness of the proposed assessment approach is that some hospital settings may be simply too busy to enable a fair attempt at the tasks, and so in these settings it may be easier to use a selected community environment. Nevertheless, this study also has a number of important strengths. First, it expands previous research with the MET by developing a novel version of the MET-HV that was both reliable and valid to predict executive problems in ABI patients with mild to moderate cognitive deficits who are still trying to improve their executive skills to successfully adjust to community living demands. In addition, our results point to the adequacy of the MET to be used with culturally different groups. Our study has the added value of using regression models to explore the predictive capacity of MET-HV indices on dysexecutive symptoms inventories, which goes beyond the correlation analyses used in previous studies. Furthermore, this is the first study that relates the results of the MET to the FrSBe, a scale that was not originally designed in tandem with the test (as is the case of the DEX) and which dissociates three functionally relevant aspects of the dysexecutive syndrome. Finally, our measures of patients’ dysexecutive symptoms are based on ratings from patients’ relatives, instead of self-report measures. This is important because patients with executive dysfunction are potentially unreliable due to lack of insight and inaccurate ratings of symptomatology (Knight et al., 2002). In fact, we found no correlations between the MET scores and the patients’ self-ratings of efficacy (pretest) and performance (posttest), which highlights the superiority of instruments such as the MET to assess patients with frontal damage in comparison to self-reports (Alderman et al., 2003).

In summary, our results indicate that everyday life executive problems can be reliably and strongly predicted by MET performance indices. This tool seems especially useful to detect executive dysfunctions often neglected by traditional neuropsychological tests and to determine the
impact of executive dysfunction on the activities of daily living.

REFERENCES


APPENDIX A

Instruction sheet given to participants

In this exercise you should complete the following three tasks:

1. **You should do the following 6 things:**
   - Collect something for the examiner from the Appointment Desk at the Outpatient Building and do what is necessary
   - Buy 3 1st class stamps
   - Buy a get well card
   - Buy a can of cola soft drink
   - Call the hospital information service using the “green user phones” and say where you are, who you are, and what time it is
   - Post something to Carlos Valls in Castellón (you can use mailboxes or suggestion boxes)

2. **You should obtain the following information and write it down in the spaces below:**
   - What is the closing time of the staff laundry on a Friday?
3. **You must meet me outside Outpatient Building 20 minutes after you have started the task and tell me the time.**

- Tell the person observing you when you have completed the exercise

4. **Whilst carrying out this exercise you must obey the following rules:**

- You must carry out all these tasks but may do so in any order
- You should spend no more than 5 euros
- You should stay within the limits of the hospital area marked by the dotted line on the map
- You should not enter any of the hospital wards or “staff only” areas
- No building should be entered other than to complete part of the task inside
- You should not go back into a building you have already been in
- You should buy no more than 2 types of items in the hospital shop
- Take as little time to complete this exercise without rushing excessively
- Do not speak to the person observing you unless this is part of the exercise

**APPENDIX B**

**Map of the area of the HRT-Granada:**
Schematic plan, list of buildings and places where the participant could perform the tasks of the test

Figure A1. E: Entrance to Main Building; Ec: Entrance to Outpatient Building; C: Community Health Center; L: Laundry Building; S: Hospital Shop; B: Hospital Cafeteria Bar; Ac: Appointment Desk; ●: Areas with bins; 1: Mailboxes or Suggestions; 2: Stamps; 3: Oreo cookies; 4: Soft drinks.