Development of a simplified version of the multiple errands test for use in hospital settings

Article in Neuropsychological Rehabilitation · June 2002
DOI: 10.1080/09602010244000039 · Source: OAI

CITATIONS
155

READS
2,167

3 authors:

Caroline Knight
The Oakleaf Group
31 PUBLICATIONS 1,069 CITATIONS

Nick Alderman
Elysium Healthcare, Borehamwood, Hertfordshire UK
76 PUBLICATIONS 4,887 CITATIONS

Paul W Burgess
University College London
147 PUBLICATIONS 18,556 CITATIONS

Some of the authors of this publication are also working on these related projects:

SASNOS-R: revision and validation of the St Andrews-Swansea Neurobehavioural Outcome Scale – an opportunity to contribute to the evolution of the second version of this established measure View project

When test scores in the normal range don’t equate to true independence: a method to convey the impact of context on ratings of neurobehavioural disability and social handicap using the ‘St Andrew’s – Swansea Neurobehavioural Outcome Scale’ (SASNOS). View project

All content following this page was uploaded by Paul W Burgess on 03 November 2015.
The user has requested enhancement of the downloaded file.
Development of a simplified version of the multiple errands test for use in hospital settings

Caroline Knight a, Nick Alderman a; Paul W. Burgess b

a St Andrew's Hospital, Northampton, UK.
b University College London, UK.

Online Publication Date: 01 June 2002

To cite this Article: Knight, Caroline, Alderman, Nick and Burgess, Paul W., (2002) ‘Development of a simplified version of the multiple errands test for use in hospital settings’, Neuropsychological Rehabilitation, 12:3, 231 - 255

To link to this article: DOI: 10.1080/09602010244000039

URL: http://dx.doi.org/10.1080/09602010244000039

FULL TERMS AND CONDITIONS OF USE

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

© Taylor and Francis 2007
Development of a simplified version of the multiple errands test for use in hospital settings

Caroline Knight and Nick Alderman

St Andrew’s Hospital, Northampton, UK

Paul W. Burgess

University College London, UK

Problems with executive functioning may have catastrophic consequences following brain injury. Valid neuropsychological assessment procedures are required if the nature and extent of these are to be understood. However, some existing measures do not adequately reflect how executive impairments are manifested in the context of everyday functioning. Shallice and Burgess (1991) described one procedure, the Multiple Errands Test (MET), which did attain this goal. While successful, it was designed for people who performed within or above the normal range when tested using existing psychometric measures: however, it is acknowledged that many patients seen in routine clinical practice perform below normal limits. Furthermore, while the procedure was carried out in a public place a range of constraints may prevent this with some patients. In this paper the utility of a simplified MET designed for use within a hospital environment is explored. Twenty neurologically healthy participants and 20 people with acquired brain injury took part. People with acquired brain injury were clearly discriminated from healthy controls through the number of errors made. Furthermore, one category of error proved highly predictive of difficulties attributable to executive dysfunction observed in the context of everyday living. While one other test correlated almost as well with these difficulties, it was argued that MET methodology is more advantageous to clinicians regarding assessment and rehabilitation.

Correspondence should be sent to Caroline Knight, Clinical Psychologist, The Kemsley Division, St Andrew’s Hospital, Billing Road, Northampton, NN1 5DG.

Email: cknigh@stanrew.co.uk

The authors would like to extend their thanks to Professor Chris Oliver, Professor Jane Riddoch, and Dr Paul Broks for the contributions they made towards this study. Thanks are also proffered to Dr Brian Moffat, Clinical Director, for allowing access to patients admitted to the Kemsley Division. Finally, the authors are grateful to all the participants who generously gave up their time through their involvement with the work reported here.
INTRODUCTION

Problems with executive functioning have been associated with catastrophic consequences for individuals with brain injury including changes in personality, behaviour, and social competency (Fogel, 1994). The need for comprehensive neuropsychological assessment to guide initial measurement, development of appropriate support systems and evaluation of rehabilitation has been highlighted (Kinsella, 1998). Quantifying executive function has historically posed a significant challenge for clinicians (Crawford, 1998) not least because lack of insight and inaccurate ratings of symptomatology are associated with impairment of these functions, thereby rendering self-report of patients potentially unreliable (Varney & Menefee, 1993).

Knowledge that performance on tests of neuropsychological functioning does somehow reflect how cognitive handicap impacts on everyday activities is obviously central to the role of the clinician. However, this relationship is not always clear or straightforward. Consider, for example, the well-cited examples of Eslinger and Damasio (1985) and Shallice and Burgess (1991) who described cases whose test performance far from adequately tallied with the range of difficulties encountered in the real world.

Lack of substantive relationships between tests of frontal lobe functioning and real life correlates has led some investigators to express the belief that such measures are not reliable predictors of everyday problems imputed to executive impairment (for example, Acimovic, Keatley, & Lemmon, 1993; Lezak, 1993; Wilson, 1993). However, more recently, Burgess et al. (1998) demonstrated that among the general neurological population such tests did correlate with observers’ ratings of behavioural manifestations of patients’ dysexecutive problems. Furthermore, executive tests were more consistent predictors of these difficulties, and with lack of insight, than tests of premorbid intellectual functions, IQ, memory, and language. Evidence was also presented which suggested that the dysexecutive syndrome may fractionate at the behavioural level into several discrete factors, and that different tests were each more or less predictive of each of these. Consequently separate tests may measure different aspects of executive functioning.

Furthermore, Shallice and Burgess (1991) previously demonstrated that patients with frontal lobe damage may be specifically impaired in everyday situations which require planning and multitasking, despite normal performance on tests of language, perception, memory and executive functioning. These investigators argued that in order to measure adequately these abilities, tests must embrace methodologies that require participants to concurrently pursue multiple tasks, over longer periods of time (than most existing tests), and without contingent feedback from the assessor. The explanation offered at that time, and subsequently (e.g., Burgess, 2000; Burgess, Veitch, Costello, & Shallice, 2000) was that those situations in which such patients had
problems were those that require subtle planning, prospective memory, and are “ill structured” (Goel & Grafman, 1997) in that there are many ways to approach the task and participants have to decide for themselves how they proceed.

Shallice and Burgess (1991) consequently described two tasks that proved sensitive to difficulties with planning and multitasking. These were the Six Element Test (SET) which required participants to organise their activities in order to carry out six tasks in a limited time period and without breaking certain rules, and the Multiple Errands Test (MET) which was carried out in a shopping precinct and required participants to buy specific items, find out certain information, to be in a particular place at a particular time, and to follow some arbitrary rules while doing these things. The methodologies of both tests incorporated those characteristics stated earlier. Shallice and Burgess subsequently described three neurological patients who performed at above average levels on measures of general ability, and whose performance on frontal lobe tests was normal, or near normal. However, in contrast, their ability to function in everyday situations was poor. Performance on the SET and MET proved consistent with everyday abilities, and was at least two standard deviations below that of neurologically healthy controls.

While the SET has received more attention to date (e.g., Burgess, 2000; Burgess et al., 2000; Levine et al., 1998; Manly, Hawkins, Evans, & Robertson, submitted) test methodologies based on the MET may also be advantageous to clinicians. For example, Alderman, Burgess, Knight, and Henman (in press) reported a simplified form of the test, that was better suited to routine clinical practice. This was validated using moderate-sized groups of neurologically healthy controls ($n = 46$) and those with acquired brain injury ($n = 50$). The test discriminated participants on the basis of two distinct error making styles, and correlated well with ratings of everyday difficulties attributable to executive dysfunction. Another reason for clinicians to use the MET is that it captures (at least in Western culture) a range of “real life” activities within the context of a “real life” environment. Observing and assessing patients’ behaviour in the real world seems an obvious thing to do if the reasons underlying such difficulties are to be fully appreciated, an activity that is not alien to other therapy professionals (e.g., occupational therapists assess cooking skills in a real kitchen). It is also the case that it is not always straightforward to relate performance on tests of executive function to rehabilitation methods or goals (Wilson, 1993). Links between assessment and rehabilitation may be more obvious when the former is undertaken within “real life” contexts.

However, not all patients are able to undergo assessment in public settings. For example, it may not be possible to test patients who have significant mobility problems (particularly wheelchair users), those with severe behavioural problems, or those detained under mental health legislation who may be prohibited from leaving hospital grounds.
Consequently, in this paper the validity of a hospital version of the MET (MET-HV) will be explored. While this version has also been simplified to enable a greater range of patients to participate, it has been specifically designed to be used in hospital environments. The intention is to produce an MET which will facilitate assessment of those patients who cannot readily be observed in public settings.

METHOD

Participants

Two groups of people were studied. The first consisted of 20 people with acquired brain injury, all of whom were patients at the Kemsley Division (a neurorehabilitation service), St Andrew’s Hospital, Northampton, UK. Participants were invited to join the study if they fulfilled the following criteria: aged 18 years or over; current full scale IQ greater than 70 (determined by performance on the WAIS-R; Wechsler, 1981); no gross perceptual problems; no impairment of dominant hand function; and intact functional language skills.

Of the 20 participants, the majority had sustained traumatic brain injury (12); the remainder had incurred damage either as a result of cerebrovascular accident (5) or a combination of the two (3). Using information in case notes, severity of brain injury was determined to be either severe or very severe for all participants. Mean time since injury was 80.9 months (SD = 62.6) and time since admission to hospital 21.2 months (SD = 18.6). The majority of participants were male (17). Age range was 20–53 years (mean = 35.6, SD = 11.3). Mean NART-R (Nelson, 1991) FSIQ was 100.8 (SD = 10.7) while current WAIS-R FSIQ was 84.4 (SD = 11.1).

In addition to completing the MET-HV, WAIS-R and NART-R, participants with acquired brain injury fulfilled a number of additional neuropsychological measures (as part of their routine clinical examination). These included tests of memory (Adult Memory and Information Processing Battery, AMIPB, Coughlan & Hollows, 1985; Rivermead Behavioural Memory Test, RBMT, Wilson, Cockburn, & Baddeley, 1985), attention and vigilance (Map Search and Visual Elevator subtests of the Test of Everyday Attention, TEA, Robertson, Ward, Ridgeway, and Nimmo-Smith, 1994) and visual perception (Visual Object and Space Perception Battery, VOSP, Warrington & James, 1991). They also completed tests reported to be sensitive to frontal lobe impairment or the presence of executive dysfunction, these being: the Behavioural Assessment of the Dysexecutive Syndrome battery (BADS; Wilson et al., 1996); a test of verbal fluency (Miller, 1984); the Cognitive Estimates Test (CET; Shallice & Evans, 1978); the Modified Card Sorting Test (MCST;
Nelson, 1976); a version of the Tower of London Test, (TOLT; Broks, et al., 1996); and versions of two procedures (hand manipulation and hand alternation) from Luria’s (1981) neuropsychological investigation, modified firstly by Christensen (1974) and further by Broks et al. (1996). In addition, the Dysexecutive Questionnaire (DEX; Burgess et al., 1996) was administered to obtain ratings of behavioural indicators of the dysexecutive syndrome. This was completed by the brain injured participants (DEX-S) as well as by a member of professional staff who knew the participant well (DEX-O). Finally, brain injured participants also rated aspects of mood using the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983).

The second group of participants comprised 20 neurologically healthy controls who were matched as closely as possible with brain injured participants regarding gender, age and NART-R FSIQ. Controls were recruited from members of staff within St Andrew’s Hospital. They were aged between 23 and 55 years (mean = 35.9, SD = 10.7). NART-R FSIQ mean was 106.9 (SD = 9.8). There was no significant difference in either age (t = 0.07, n.s.) or NART-R FSIQ (t = 1.86, n.s.) between the two groups. As well as completing the MET-HV and NART-R, control participants also completed the DEX-S and the HADS.

**Test Environment**

The MET-HV was conducted within the grounds of St Andrew’s Hospital, an independent psychiatric hospital located centrally within the UK. That part of the site employed for the purpose of the test is shown in Figure 1.

**Description of the MET-HV**

This comprised a simplified version of the procedure described by Shallice and Burgess (1991). Three principal modifications were made which mirrored those described by Alderman et al. (in press) in that: clarity was enhanced through provision of more concrete rules than those contained in the original MET; task demands were simplified; and participants were given an instruction sheet which explicitly directed them to record designated information. The goal of these modifications was to simplify the MET sufficiently to enable a wider range of participants to be tested in contrast to the patients with superior IQ described by Shallice and Burgess (1991) for whom the test was originally devised. The sheet containing a summary of the test given to participants is reproduced in Appendix 1, while the instructions read to them by an assessor are shown in Appendix 2.

**Tasks.** Participants were required to achieve four sets of simple tasks, totalling 12 separate subtasks in all. The first required participants to attain six specific goals which included purchasing three items, collecting an envelope
Figure 1. Schematic plan view of that area of St Andrew’s Hospital used for the test. Key: dark shaded areas = buildings; unshaded areas = roads; “E” = main entrance; “L” = staff library; “S” = hospital shop; “K” = Kemsley reception; “•” = car park. Numbers refer to locations where tasks can be accomplished: post box (1); stamps (2); Mars Bar (3); Coca Cola (4); get well card (5).
from reception, using the internal telephone, and posting something to an external address (see Appendix 1). The second involved obtaining and writing down four items of designated information (for example, the opening time of the hospital shop on a Saturday; see Appendix 1). In the third task, participants were required to meet the assessor outside the hospital reception 20 minutes after the test had begun and state the time. The final task was that the participant was required to inform the assessor when they had finished the test.

Rules. To reduce ambiguity and help simplify task demands, the number of rules was expanded from the six described by Shallice and Burgess (1991) to nine. These were essentially the same as those described by Alderman et al. (in press). They were also made more explicit than those contained in the original MET, and were clearly stated on the instruction sheet carried throughout the test by the participants (see Appendix 1).

Procedure

The test began outside the main hospital reception (see Figure 1) where participants were briefed regarding what was required of them by an assessor. First, separate ratings were obtained regarding the following two statements: “How efficient would you say you were with tasks like shopping, finding out information, and meeting people on time?”; and “How well would you say you know the hospital grounds?”. The goal was to quantify participants’ perceptions of their ability to carry out the types of task undertaken in the test, and to obtain a measure of familiarity with the hospital environment. The efficiency item was rated using a 10-point Likert type scale with weighted end points (1 = hopeless, 10 = excellent); familiarity was measured using a four-point scale (0 = not at all, 1 = somewhat, 2 = fairly well, 3 = very well).

Participants were given a copy of the test instructions (see Appendix 1) on a clipboard, a pen, a carrier bag, and a £5 note. A wrist watch was also given to any participant who did not have one. The assessor then read aloud the instructions (see Appendix 2). The rules were explained with reference to the instruction sheet given to participants. It was emphasised that the assessor would follow and observe participants at a distance and should not be spoken to unless this was a part of the test. Participants were then prompted to ask any questions they had and then to summarise what they were expected to do. When required, the assessor repeated instructions until she was satisfied each participant was familiar with the demands of the test. At this point, the assessor began the test with the statement “Begin the exercise”.

Once the test was under way, each participant was followed at a distance by the assessor who made written notes for the purpose of recording behaviour and performance.
At the end of the test each participant was prompted to rate the question “How well do you think you did with the task?” using a 10-point scale as before (1 = hopeless, 10 = excellent). They were then debriefed regarding their performance.

**Analysis of individual participant performance**

After the test, the assessor used her notes to determine error scores regarding MET-HV performance. Errors were categorised using the same definitions proffered by Shallice and Burgess (1991), namely: (1) *inefficiencies*—where a more effective strategy could have been applied; (2) *rule breaks*—where a specific rule (social, or one of the nine explicitly defined within the test) was broken; (3) *interpretation failure*—where the requirements of a task had been misunderstood; (4) *task failures*—where any one of the 12 tasks had not been fully completed. An attempt was also made to quantify the range of strategies employed (for example, the number of times participants asked other people for help).

A concern was that as the MET-HV was a less structured test than most psychometric procedures inter-rater reliability may be low (Crawford, 1998). Consequently, errors were scored separately by two independent raters: first, by the assessor who accompanied the participant; and second, the consultant neuropsychologist. In all cases, error scores generated by the assessor were used for the purpose of analysis.

**RESULTS**

Reliability of the MET-HV

*Inter-rater reliability*

The method proposed by Everitt (1996) was used to calculate intraclass correlation coefficients for each category of error, using the two sets of scores for each participant. Inter-rater reliability proved good (see criteria suggested by Spitzer, Fleiss, & Endicott, 1978), ranging from .81 (interpretation failures) to 1.00 (rule breaks).

*Internal consistency*

Cronbach’s alpha was used to test the reliability of each item of the MET-HV to predict total error score. Internal consistency proved satisfactory at .77 (see Loewenthal, 1996).
Comparisons between brain injured and control participants’ MET-HV performance

Summary statistics regarding the different error types are shown in Table 1. Possible differences between group means were investigated through application of $t$-tests. Separate variance estimates were used for those comparisons where Levine’s test indicated this was appropriate (see Howell, 1997, pp. 198–199). It can be seen that brain injured participants made significantly more rule breaks and total errors than controls, and achieved significantly fewer tasks. Performance on the test proved to discriminate well between people with and without brain lesions. A cut-off score of total errors made (7) was determined using the top 5th percentile of that of controls. Using this, 85% of brain injured participants were correctly identified as such through their MET-HV performance (only one control would have been misclassified using this criteria).

### TABLE 1

Differences in MET-HV performance between control and brain injured participants

<table>
<thead>
<tr>
<th>MET-HV Errors</th>
<th>Controls</th>
<th>Brain injured</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inefficiencies</td>
<td>0.25</td>
<td>0.30</td>
<td>0.26</td>
<td>.796</td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td>(0.73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpretation failures</td>
<td>0.35</td>
<td>0.55</td>
<td>0.93</td>
<td>.357</td>
</tr>
<tr>
<td></td>
<td>(0.59)</td>
<td>(0.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rule breaks</td>
<td>1.50</td>
<td>6.70</td>
<td>3.52</td>
<td>.002*</td>
</tr>
<tr>
<td></td>
<td>(1.28)</td>
<td>(6.47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task failures</td>
<td>1.70</td>
<td>5.75</td>
<td>6.82</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td></td>
<td>(1.08)</td>
<td>(2.43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total errors</td>
<td>3.80</td>
<td>13.30</td>
<td>5.25</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td></td>
<td>(1.64)</td>
<td>(7.92)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategies</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency looked at map</td>
<td>.065</td>
<td>.10</td>
<td>2.83</td>
<td>.008*</td>
</tr>
<tr>
<td></td>
<td>(0.75)</td>
<td>(0.48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency read signs</td>
<td>1.60</td>
<td>0.85</td>
<td>2.90</td>
<td>.006*</td>
</tr>
<tr>
<td></td>
<td>(0.68)</td>
<td>(0.93)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial planning time (seconds)</td>
<td>72.55</td>
<td>24.15</td>
<td>1.85</td>
<td>.720</td>
</tr>
<tr>
<td></td>
<td>(105.64)</td>
<td>(50.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency carried out multiple tasks</td>
<td>0.35</td>
<td>0.05</td>
<td>2.49</td>
<td>.019*</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td>(0.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency requested help</td>
<td>2.40</td>
<td>3.25</td>
<td>1.15</td>
<td>.256</td>
</tr>
<tr>
<td></td>
<td>(2.42)</td>
<td>(2.25)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Separate variance estimates used.
Other differences between groups were also noted. Five different strategies used by participants were identified and are shown in Table 1. In comparison to brain injured participants, controls were found to look at maps and signs more frequently, and pursue two or more tasks concurrently. While there was a tendency for controls to ask others for help less often and to spend longer planning before beginning, these did not prove significant.

However, use of these strategies had little impact on MET-HV performance. Among controls, the only significant relationship between frequency of strategy use and errors was a negative correlation between simultaneous pursuit of two or more concurrent tasks and fewer interpretation failures (–.45, \( p = .047 \)). Similarly, among people with acquired brain injury, making more frequent use of signs was associated with fewer task failures (–.46, \( p = .042 \)).

**Effects of other variables on the MET-HV**

**Age**

There were no significant correlations between age and MET-HV errors for controls. However, among brain injured participants, age correlated significantly with total errors (.65, \( p = .002 \)), rule breaks (.57, \( p = .008 \)), and task failures (.69, \( p = .001 \)).

**FSIQ**

There were no significant correlations between NART-R FSIQ and MET-HV errors for either the brain injured or control group, with the exception that a higher NART-R FSIQ score was associated with fewer task failures for controls (–.47, \( p = .039 \)). Furthermore, there were no significant correlations between WAIS-R FSIQ and MET-HV errors for the brain injured group.

**Familiarity**

Both groups of participants most frequently rated the hospital grounds as “somewhat” or “fairly well-known” (brain injured = 85%; controls = 75%). There was no significant difference between means regarding ratings of familiarity between groups (\( t = 1.46, \text{n.s.} \)).

The possible contribution of familiarity to MET-HV performance was examined. There were no significant correlations between familiarity and errors for either group, except that higher familiarity ratings were associated with fewer task failures for brain injured participants (–.66, \( p = .001 \)). Furthermore, there were no correlations found regarding familiarity ratings and use of those strategies shown in Table 1.
Competency ratings

There was little difference between ratings of competency regarding tasks analogous to the MET-HV made by either brain injured participants (mean = 7.50, SD = 2.04) or controls (mean = 7.45, SD = 1.54; \( t = 0.09, \text{n.s.} \)). Furthermore, despite significant differences between groups regarding errors and strategies used, there was little difference between ratings of how participants thought they had performed on the MET-HV (brain injured participants mean = 6.40, SD = 2.93; control participants mean = 7.45, SD = 1.23; \( t = 1.48, \text{n.s.} \)). Most participants rated themselves above the midpoint on both scales. Despite the optimism of both groups, there were no correlations between either sets of ratings and actual performance on the test. The only exception was the association between competency ratings with tasks like the MET-HV and rule breaks (–.47, \( p = .037 \)) among controls. However, when the significance level was adjusted to reduce risk of Type 1 error because multiple correlations were considered using the Bonferroni method, this too disappeared.

Mood

Brain injured participants rated themselves as more anxious (mean = 6.85, SD = 3.10) than controls (mean = 4.45, SD = 2.61; \( t = 2.65, \ p = .012 \)), and also more depressed (mean = 6.50, SD = 3.55; controls mean = 2.40, SD = 2.16; \( t = 4.41, \ p < .001 \)). Despite these differences, no significant correlations were found between the HADS and errors on the MET-HV for either group. Also, it should be noted that the mean score for anxiety and depression for both groups fell within the “normal” range.

Cognitive functioning

Finally, performance on tests of general ability (WAIS-R FSIQ), memory (AMIPB, RBMT), and visual perception (VOSP) were examined to determine their relationships with the MET-HV. Scores for each component comprising these tests were correlated in turn with the five error categories from the MET-HV (inefficiencies, interpretation failures, rule breaks, task failures, and total errors). Performance on some test items did correlate significantly with one or more of the MET-HV error categories (for example, Dot Counting from the VOSP with rule breaks, –.54, \( p = .014 \)). However, when significance levels were adjusted using the Bonferroni method because multiple correlations were considered, only the relationship between the Profile Score from the RBMT and the number of task failures retained significance (–.57, \( p < .05 \)). (It should be noted that when sets of multiple correlations feature in the analysis, such as those considered under the current subheading, this method is used to reduce the risk of Type 1 error among those considered for each separate
investigation; for example, the relationship between tests of cognitive functioning and the MET-HV.)

Validity

While MET-HV performance differed between brain injured participants and controls, a particular function of the test is to reflect and confirm the presence and severity of everyday difficulties attributable to executive dysfunction. The relationship with existing measures of executive functioning will now be examined.

Relationship with traditional frontal lobe tests

First, possible associations between the MET-HV and existing tests of frontal lobe function were considered (i.e., verbal fluency, CET, MCST, TOLT, and hand manipulation/hand alternation). The only significant correlations were between MCST percentage perseverative errors with rule breaks (.66, \( p = .002 \)), task failures (.49, \( p = .027 \)), and total MET-HV errors (.67, \( p = .001 \)). When the Bonferroni adjustment was applied, only the correlations between perseverative errors with rule breaks and total errors maintained significance (\( p < .05 \)).

Relationship with ecologically sensitive executive tests

Next, associations between the MET-HV and those tests that claim to be rich in ecological validity were investigated (i.e., BADS and Visual Elevator/Map Search from the TEA). Significant correlations were found between the BADS Profile Score with rule breaks (–.51, \( p = .022 \)), task failures (–.58, \( p = .007 \)), interpretation failures (.64, \( p = .003 \)), and total MET-HV errors (–.57, \( p = .009 \)). Regarding the two subtests administered from the TEA, only Visual Elevator correlated with the MET-HV (accuracy scaled score with rule breaks, –.47, \( p = .037 \), task failures, –.47, \( p = .047 \), total errors, –.49, \( p = .028 \), and timing scaled score with interpretation failures, .45, \( p = .047 \)). After significance levels were adjusted for multiple comparisons, significant correlations between the BADS profile score and MET-HV task failures, interpretation failures and total errors were preserved (\( p < .05 \)).

Relationship with DEX questionnaire

Finally, the relationship between the MET-HV and ratings regarding behavioural indicators of the dysexecutive syndrome was examined. Consistent with findings reported elsewhere (Wilson et al., 1996; Burgess et al., 1998), brain injured participants’ ratings on the DEX were more variable (mean = 22.10, SD = 13.67) than neurologically healthy controls (mean = 14.40, SD = 5.62; \( t = 2.33, p = .028 \)). In addition, ratings made by
rehabilitation staff concerning brain injured participants were higher than those made by this group (mean staff rating = 37.10, SD = 12.63; t = 3.78, p = .001). Brain injured participants perceived themselves as having fewer and less severe functional problems than those people who knew them well.

In order to determine whether performance on the MET-HV was predictive of behavioural indicators of the dysexecutive syndrome a series of Pearson correlations were computed between the five error categories with the DEX. It is recognised that lack of awareness and reduced insight may be characteristic of acquired brain injury (Sazbon & Gros Wasser, 1991). Indeed, this is supported by the finding here that patients underestimated the extent of their difficulties. As a consequence, ratings made by others about patients, rather than the ratings patients made about themselves have been used in previous studies to determine the relationship of a range of frontal lobe and executive tests with the dysexecutive syndrome (Wilson et al., 1996; Burgess et al., 1998). This methodology was used in the present study.

The total sum of ratings made in response to the 20 items comprising the DEX provided an overall estimate regarding both the presence and severity of behavioural indicators of the dysexecutive syndrome. However, there is evidence to suggest that this fractionates at the level of everyday behaviour. Burgess et al. (1998) subjected others’ DEX ratings to a varimax rotated factor analysis and found five factors, namely: (1) inhibition—behavioural manifestations of disinhibition or inability to inhibit a habitual response; (2) intentionality—difficulties in formulating goal-oriented plans and executing these satisfactorily; (3) executive memory—disturbances of memory associated with executive dysfunction, especially those of confabulation and perseveration; (4) positive affect—positive emotional and personality changes associated with executive dysfunction, including aggression and variable motivation, and; (5) negative affect—negative emotional and personality changes, including apathy and shallow affect.

Consequently, total DEX-O scores and the sum of ratings pertaining to each of these five factors were correlated with the error scores attained on the MET-HV. Table 2 shows significant associations were found between two MET-HV error categories with two of the five DEX-O factors, and with sum of DEX-O ratings. Failure to achieve set tasks on the MET-HV was positively correlated with the intentionality and executive memory factors, and with total DEX-O rating. In addition, interpretation failures were negatively correlated with total DEX-O rating. When significance levels are adjusted for multiple comparisons, only the association between intentionality and task failures is preserved. Nevertheless, the overall pattern of association between the MET-HV as a test procedure and individual behavioural measures is encouraging.

Table 2 shows that the principal relationships between the MET-HV and the DEX questionnaire fall within the domain of failure to achieve set tasks. However, it will be recalled that there was a tendency for older brain injured
participants to attain less tasks, break more rules, and make more errors generally. Furthermore, poor memory (as measured by the RBMT) was also associated with a tendency to achieve less tasks, while increased familiarity with the hospital environment was related to less errors of this type. In order to determine whether performance on the MET-HV reliably reflected the presence of executive dysfunction, as opposed to any mixed effects from these three confounding variables, relationships between errors and the DEX-O were re-evaluated. Partial correlations were computed which controlled the effect of age regarding rule breaks and total errors; and age, familiarity and memory function with respect to task failures. These are also shown in Table 2. Controlling the effect of age did not change the previous findings regarding

<table>
<thead>
<tr>
<th>MET-HV correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
</tr>
<tr>
<td>Task failures</td>
</tr>
<tr>
<td>Factor 2: intentionality</td>
</tr>
<tr>
<td>Factor 3: executive memory</td>
</tr>
<tr>
<td>Total DEX-O ratings</td>
</tr>
<tr>
<td>Interpretation failures</td>
</tr>
<tr>
<td>Total DEX-O ratings</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Partial correlations*</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
</tr>
<tr>
<td>Task failures</td>
</tr>
<tr>
<td>Factor 1: inhibition</td>
</tr>
<tr>
<td>Factor 2: intentionality</td>
</tr>
<tr>
<td>Factor 3: executive memory</td>
</tr>
<tr>
<td>Factor 4: positive affect</td>
</tr>
<tr>
<td>Total DEX-O ratings</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other tests correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive estimation</td>
</tr>
<tr>
<td>Factor 1: inhibition</td>
</tr>
<tr>
<td>Factor 2: intentionality</td>
</tr>
<tr>
<td>MCST perseverative errors</td>
</tr>
<tr>
<td>Factor 2: intentionality</td>
</tr>
<tr>
<td>BADS: zoo map</td>
</tr>
<tr>
<td>Factor 1: inhibition</td>
</tr>
<tr>
<td>Factor 2: intentionality</td>
</tr>
<tr>
<td>Factor 3: executive memory</td>
</tr>
<tr>
<td>Factor 4: positive affect</td>
</tr>
<tr>
<td>Total DEX-O ratings</td>
</tr>
</tbody>
</table>

*age partialled out for rule breaks and total errors; age, RBMT profile score and familiarity with hospital grounds partialled out for task failures
either rule breaks or total MET-HV errors. However, rather than weakening the relationship between the test procedure and ratings of executive dysfunction, Table 2 shows that partialling out the influence of the compounding variables increased it. Failure to achieve set tasks was now significantly correlated with five out of the six DEX-O measures. Clinically, this is very encouraging. The strength of these associations also improved: Even when the Bonferroni correction for multiple correlations was applied, the strength of the relationships between task failures with inhibition, intentionality, executive memory and sum of DEX-O ratings remained significant. These results suggest that performance on the MET-HV both reflects, and is indicative of, behavioural indicators of executive dysfunction (as measured by the DEX).

However, how well does the MET-HV compare to other measures in the prediction of executive impairment? To help answer this question, scores from the other frontal lobe and executive tests were also correlated with the DEX-O. Those which attained individual significance are also shown in Table 2. Cognitive estimation was correlated with the inhibition and intentionality factors, while MCST perseverative errors were also associated with intentionality. While several of the tests comprising the BADS correlated with DEX-O, the best and most consistent coefficients were achieved with the Zoo Map task. Table 2 shows these were the same items which attained significance regarding MET-HV task failures using partial correlations. When the Bonferroni procedure was applied, neither CET nor MCST perseverative errors remain significant, while it was retained for Zoo Map with the inhibition and executive memory factors, and sum of DEX-O ratings.

The moderate correlation between Zoo Map and MET-HV task failures (−.52, \( p = .018 \)) suggests both tests may be measuring broadly similar but not identical processes. Given they both correlated well with DEX-O ratings, which of these measures provides the best prediction of behavioural indicators of executive dysfunction? Table 2 shows that Zoo Map correlates with five of the six sets of DEX-O scores, while task failures were significantly associated with three. The two tests were each correlated with the intentionality and executive memory factors, and the total sum of DEX-O ratings. Williams’ (1959) procedure was used to test for significant differences between the correlations to determine which, if either, test best predicted observational ratings of executive dysfunction (see Howell, 1997, pp. 264–265). Analysis confirmed the correlations between Zoo Map with executive memory (\( t = 3.15, \ p = .003 \)) and sum of DEX-O ratings (\( t = 3.55, \ p = .002 \)) were significantly better than those regarding MET-HV task failures. However, task failures proved to be significantly more associated with the intentionality factor (\( t = 3.64, \ p = .002 \)).

However, Table 2 reminds us that when the compounding effects of age, memory and familiarity are controlled, associations between task failures and DEX-O ratings increase to the extent that reasonable correlations are achieved with all but one of the DEX factors, as well as the overall sum of ratings.
Possible differences between these correlations regarding Zoo Map and task failures were not undertaken because partial correlations were computed for the former but not the latter.

**DISCUSSION**

In this paper a hospital version of the Multiple Errands Test was described, its psychometric properties examined, and its role in clinical assessment investigated.

Agreement between raters regarding performance errors was shown to be good: Scoring was straightforward and proved highly reliable. Internal consistency results suggested that the majority of the items were useful in contributing to the overall score.

People with acquired brain injury were found to make significantly more errors than neurologically healthy controls. However, performance within the brain injury group was more variable, which suggested that not all people with neurological damage made errors outside the expected range. In fact, when this hypothesis is examined the test proved to be able to discriminate between lesion and non-lesion groups with a very high degree of accuracy: Using a cut-off of 7 or more errors (i.e., 5th percentile of controls) 85% of brain injured people were correctly classified as such, while only one control would have been incorrectly assigned to the lesion group. This compares very favourably to existing tests of frontal lobe and executive functioning, some of which have been found to lack test-sensitivity in discriminating controls from patients (see Burgess et al., 1998; Alderman et al., in press).

While neurologically healthy controls made fewer errors in comparison to participants with acquired brain injury, reasons to account for this from strategies participants used were less clear. While controls consulted a map of the hospital more frequently, used signs more often, and were better able to pursue two or more tasks concurrently, these strategies had little obvious benefit on performance. However, a potentially important finding was that whilst brain injured participants consulted signs less often than controls, those that used them more frequently exhibited a tendency to fail fewer of the set tasks. We will return to the use of this strategy later.

Qualitatively, there were differences in performance between the two groups. It was evident that the brain injured group were more impulsive, as some of them began the test almost immediately without considering any plan beforehand. They also broke rules more frequently. As a group, they resorted to use of environmental cues less often, and were more reliant on others, for example, as indicated by the greater number of times they asked the assessor for help. In addition, the “knowing and doing” dissociation was often observed (Teuber, 1964), for example, a strategy to attain a task was verbalised but then not followed. Interestingly, the goal of simplifying the test does appear to have
been achieved as both groups made few interpretation failures. However, despite understanding the demands of the test, brain injured participants nevertheless went on to make significantly more errors than controls.

Measures of general ability (WAIS-R, NART-R) were not predictive of MET-HV performance among people in either group. However, there was a tendency among older brain injured participants to achieve fewer of the set tasks. Memory impairment, as measured by the RBMT, was also associated with more errors of this type. Conversely, those patients who rated themselves as being more familiar with the hospital environment were inclined to achieve more of the tasks successfully.

It would have been surprising if age and memory effects had not been found. There is evidence of age-related decline in frontal systems and on unstructured tasks reliant on supervisory processes (Levine et al., 1998), particularly following brain injury (Burgess & Shallice, 1996). Aspects of memory, particularly working and prospective memory, have been closely related to executive functioning in that performance of previously formed intentions rely on the intact functioning of the supervisory attentional system or central executive which in turn may use prior knowledge in planning (Baddeley, 1986; Kinsella, 1998). The RBMT consists of several tasks involving prospective memory and the particularly strong relationship with this measure supported this.

Despite being unable to quantify what proportion of MET-HV errors made by brain injured participants was due to memory impairment in comparison to controls, the principal goal of the study was achieved. Performance on the MET-HV not only reliably discriminated lesion from non-lesion cases, it also proved to be a good predictor of behavioural indicators of executive impairment that are evident in the course of everyday function. As a consequence, it joins the ranks of other recent test procedures which claim to have good “ecological validity”. From the participants’ point of view it also has excellent face validity. Error making behaviour correlated highly with ratings on the DEX questionnaire made by others regarding brain injured participants. However, it is of note that this was limited to only one of the five error categories, task failures. This behaviour was itself partly a function of age, memory and familiarity with the hospital environment. Nevertheless, when the possible confounding effects of these variables was statistically controlled, failure to achieve set tasks remained strongly associated with the inhibition, intentionality, and executive memory factors of the DEX, and with the total sum of ratings. This certainly suggests the possibility that this aspect of MET-HV performance is measuring specific aspects of cognitive function.

However, it is equally notable that relationships between the other three error categories, and total errors made, with DEX-O ratings were lacking. Only two sets of associations were found. The negative correlation between interpretation failures and sum of DEX-O ratings cannot be readily explained. The second is more in line with expectations in that rule breaking behaviour
observed during the test was positively correlated with Question 14 on the DEX that rates perseveration (.52, \( p = .018 \)). However, while clinically interesting, both these correlations lose significance under conditions of Bonferroni correction. The lack of significant relationships between rule breaking errors and behavioural indicators of a dysexecutive syndrome in particular is surprising, as disinhibition and disregard for social and other norms are often cited as key indicators of frontal lobe damage.

In a parallel version of the MET-HV which uses a virtually identical methodology, but is conducted within the environment of a shopping centre, Alderman et al. (in press) found that breaking rules and DEX-O ratings did indeed correlate. However, this study used a larger sample size (50 brain injured participants) within which two distinct error-making patterns were identified. Patients either predominantly failed to achieve the set tasks or displayed a tendency to break rules. Meaningful associations between test performance and DEX ratings only became evident when patients were assigned to sub-groups whose membership was determined by error style. Unfortunately, the smaller number of patients in the present study did not permit creation of meaningful sub-groups. Another reason why rule breaking behaviour was found to be associated with the DEX in the Alderman et al. study was that a system of weighted error scores was used. The MET performance of a moderate-sized group of neurologically healthy controls (\( n = 46 \)) was examined in detail: When the two groups were compared it was evident that not only did brain injured participants make more errors, but also there were striking qualitative differences regarding these. Indeed, while some errors were typical of both groups, many were characteristic of participants with neurological damage. When error scores were subsequently weighted to reflect these important qualitative differences, correlations with the DEX-O also changed. Because of the smaller number of participants studied here, it was not possible to create a valid means of weighting scores pertaining to the hospital environment. If this had been feasible, robust associations between rule breaking behaviour and ratings of executive dysfunction may have been more evident.

A third reason to account for poor associations between breaking rules and DEX-O ratings may be due to differences between the two environments in which the parallel versions of the MET took place. Experience of both environments suggests that the shopping centre proved a more distracting setting (many people, noisier, many complex and competing visual stimuli) than that of the hospital (fewer people, comparatively quiet, more consistent visual stimuli). One consequence of this may be there was more opportunity to exhibit rule breaking behaviour in the shopping centre. Additional differences regarding cue saliency may also have been important. While being less distracting, the hospital also had more signs and maps on view. In the Alderman et al. study, it was suggested that difficulties with monitoring accounted in part
for rule breaking behaviour. It may be reasonable to conclude that in a less distracting environment in which cue saliency regarding test goals is increased through availability of signs and maps, difficulties with monitoring are reduced, and with it the behavioural manifestations of this difficulty. This idea is partly supported by the finding that use of signs and rule breaking behaviour among the brain injured group was negatively correlated (−.36).

Only task failures were found to be predictive of behavioural ratings of problems attributable to executive dysfunction; however, the strength of these correlations were high so it may be reasonably concluded that performance on the MET-HV as a whole provided valid and reliable information about executive functioning. The only other test administered here which was similarly correlated with the DEX was Zoo Map from the BADS. Both MET-HV task failures and Zoo Map correlated well with the inhibition, intentionality, executive memory, and positive affect factors from the DEX, and the total sum of DEX ratings.

The finding that both tests relate similarly to the DEX should probably come as no surprise as on the face of it both tests have similar task demands in that: they have a planning requirement; participants must “visit” various locations and go from place to place in a sensible fashion; and they must adhere to a set of arbitrary rules. It has been argued that the MET has a strong multitasking requirement, and it is this that distinguishes it from most other tests of executive functioning (Burgess, 2000; Burgess et al., 2000; Shallice & Burgess, 1991). Furthermore, multitasking is thought to have three central components: planning, retrospective and prospective memory. Given the obvious similarities between the tests it is very likely that Zoo Map also shares similar planning and memory demands to the MET-HV.

The obvious difference between the two tests is that one is “real life” whereas the other is not. However, given that they probably both tap similar cognitive processes and that they share similar relationships with everyday behavioural symptoms of the dysexecutive syndrome is very encouraging, and suggests both possess good construct validity.

While there is a degree of equivalence between the tests it is probably unlikely that both measure exactly the same range of cognitive processes: indeed, the strength of the correlation between the two tests suggests that while both probably measure broadly similar facets of executive functioning these may not be identical. This was borne out by subsequent analysis which suggested Zoo Map was a better measure of executive memory, whilst MET-HV task failures offered superior estimation of intentionality. One conclusion from this result is that both tests should be undertaken to help build the most complete clinical picture.

A possible weakness of the MET-HV is that reliability is probably more diluted than Zoo Map, as less control over the testing environment is available. Of course, this is true, and for that reason it must have poorer reliability than
traditional psychometric measures. However, while test–retest has not been specifically investigated here, data from other studies suggests that despite greater environmental variability, reliability of MET methodologies is surprisingly robust. For example Knight (1999) compared the performance of 10 neurologically healthy participants who were tested twice using the MET-HV: once in the grounds of St Andrew’s Hospital (as reported here) and again within the confines of the local general (medical) hospital. A counterbalanced design was used. A high correlation (.83) between overall error scores from each of the two hospital sites was found. Further support regarding reliability comes from the simplified version of the MET described by Alderman et al. (in press): While the finding was not reported within their study, a mixed group of 20 lesion and control participants were tested twice, one week apart, at the shopping centre used for their version. While it might reasonably be suggested that such an environment may be even less predictable than that of a hospital, a correlation of .81 was achieved between the two sets of overall error scores. Both these results suggest that reliability of MET methodologies is reasonably robust and perhaps less susceptible than might be thought to environmental variability and therefore may be relied on with some confidence.

Participation in the MET-HV may also yield different clinical information that is not obvious through performance on other psychometric tests. For example, it involves free interactions and the opportunity to observe social behaviour in the real world. Performance may also provide a clearer and more seamless link with rehabilitation. For example, this study suggests that people who made greater use of signs were likely to achieve more of the set tasks. An obvious implication of this is that executive functioning may be improved by training people to make use of existing cues in the environment, or to circumvent difficulties with planning through use of some aid. For example, Alderman, Knight, Rutterford, and Swan (2000) identified problems with planning and initiation in three people with acquired brain injury using the parallel shopping centre version of the MET; task success was subsequently enhanced by repeating the MET and giving these patients the instructions for the test in the form of a written plan they were instructed to follow.

The MET-HV was designed so that its methodology should be applicable in most hospital settings: Tasks were selected that should be possible to undertake given the range of resources likely to be found in such environments (for example, most hospitals have a shop and car parks). However, it should be noted that two of the tasks need to be adapted to the individual location: the name and address of the person to “post something to”; and the ward to contact by telephone (see Appendix 1).

However, it is recognised that constraints on time and resources mean that not all clinicians will be able to test their patients using a multiple errands methodology. When this is the case, the similarities in format, cognitive demands, and relationship with behavioural symptoms of executive disorder,
all suggest that Zoo Map will provide clinicians with a comparable measure of multitasking ability.

In conclusion, this preliminary investigation regarding the utility of a multitasking assessment tool that is simple and robust enough to use in routine clinical practice with a wide range of people, within a hospital setting, looks very encouraging. The methodology of the original MET described by Shallice and Burgess (1991) has been retained in that the test necessitates concurrent pursuit of multiple tasks, over longer periods of time, and without feedback from the assessor. This version of the MET has been designed to enable assessment of those individuals in whom, for various reasons, appraisal in more public places is either problematic or not an option. Only one of the five possible error scores was found to be predictive of behavioural indicators of the dysexecutive syndrome: however, this proved highly predictive. Only correlations with one other test (of those administered here) proved compatible. It is also possible to make qualitative observations about patients in real-life settings that cannot be gleaned when people are tested in the detached environment of an office or ward. Furthermore, performance on the MET may make choice of rehabilitation goals clearer.

The findings reported in this paper and elsewhere confirm that test procedures that claim to have ecological validity have much to offer the clinician. It appears MET methodology is particularly adept at measuring how and to what extent executive impairment impacts at the level of everyday functioning. Consequently, development of this means of assessment, generalisation of the procedure described here to other hospital environments, and its role in rehabilitation should be encouraged.

REFERENCES


Manuscript received June 2001
Revised manuscript received November 2001
APPENDIX 1:
Instruction sheet given to participants

INSTRUCTIONS

In this exercise you should complete the following three tasks:

1. You should do the following 6 things:
   - Collect something for the examiner from Main Reception and do what is necessary
   - Buy 4 1st class stamps
   - Buy a get well card
   - Buy a can of Coca-Cola
   - Telephone Kemsley Reception and say where you are, who you are, and what time it is
   - Post something to Caroline Knight in Birmingham

2. You should obtain the following information and write it down in the spaces below:

   1. What is the closing time of the staff library on a Friday?
   2. What is the opening time of the hospital shop on a Saturday?
   3. What is the price of a Mars Bar?
   4. How many public carparks are there in the hospital grounds (not including staff or disabled only parking)?

3. You must meet me outside Main Reception 20 minutes after you have started the task and tell me the time.

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

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 £2.50
- You should stay within the limits of the hospital grounds
- 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 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

Your examiner was:
Caroline Knight
University of Birmingham, School of Psychology, Edgbaston, Birmingham, B15 2TT.
Appendix 2: Instructions for Participants

This test makes use of the following items:

- Pen/pencil
- Instructions on a clipboard for the participant
- carrier bag
- £5 note
- Examiner to have stopwatch, pad, and pen to record observations of the participant

Ensure the participant is wearing a watch, and that the envelope (marked for the urgent attention of Caroline Knight) is left at Main Reception before beginning the procedure.

Before starting, obtain ratings for efficiency and familiarity (see separate sheet).

On completion of the task, obtain the rating for how well the participant felt he or she executed the task (see separate sheet).

Begin the task outside the main reception of the hospital. Give the participant the clipboard, pen/pencil, carrier bag and £5 note. Read the following instructions to the participant:

“In this exercise I want you to complete 3 tasks. The tasks are: to do the 6 things listed on this sheet (examiner to indicate and describe items on sheet); to obtain and write down 4 pieces of information (examiner to indicate and describe items on sheet); and to meet me here outside main reception 20 minutes after I have said “. . . begin the exercise” and tell me the time.

However, while completing this exercise you must obey the rules listed on your instruction sheet (examiner to indicate and describe rules on sheet).

You must carry out all of the tasks but you may do so in any order. You should spend no more than £2.50: although I have given you £5 you should spend no more than £2.50. You should stay within the limits of the hospital grounds. This means you must not leave by any of the entrances/exits. You must not enter any ward area or area where staff only are allowed. No building should be entered other than to carry out part of the task, so if you go into a building it must be with the intention of completing part of the task.

You should not go back into a building you have already been in, so if you have been into a particular building you should not go back into it again. You should buy no more than 2 items from the hospital shop. Take as little time as possible to complete this exercise without rushing excessively. There is no time limit to completing the task.

During the exercise I shall be following you from a distance and observing what you are doing. Please do not speak to me unless this is part of the exercise.

Finally, approach me and tell me when you have completed the exercise.

Is that clear, have you any questions? (Clarify any questions the participant has.)

Now tell me what you must do. (Ensure participant is clear about what they must do.)

Begin the exercise. (Examiner to start timing at this point.)