Ecological validity of a simplified version of the multiple errands shopping test

NICK ALDERMAN,[1] PAUL W. BURGESS,[2] CAROLINE KNIGHT,[1] AND COLLETTE HENMAN[1]

[1] The Kemsley Division, St Andrew's Hospital, Billing Road, Northampton NN1 5DG, UK
[2] Institute of Cognitive Neuroscience, UCL (University College London), 17 Queen Square, London WC1N 3AR, UK

Abstract

Shallice and Burgess (1991) reported the utility of the Multiple Errands Test (MET) in discriminating executive deficits in three frontal lobe patients with preserved high IQ, who were otherwise unimpaired on tests of executive function. The aim of this study was to ascertain the value of a simplified version of the MET (MET–SV) for use with the range of people more routinely encountered in clinical practice. Main findings were as follows: 1) The test discriminated well between neurological patients and controls, and the group effects remained when the difference in current general cognitive functions (WAIS–R FSIQ) was taken into account. 2) The best predictors of performance in the healthy control group ($n = 46$) were age and the number of times participants asked for help (with more requests associated with poorer performance). 3) In the neurological group, two clear patterns of failure emerged, with performance either characterized by rule breaking or failure to achieve tasks. These two patterns were associated with different dysexecutive symptoms in everyday life. 4) The patients not only made more errors than controls, but also different ones. A scoring method that took this into account markedly increased test sensitivity. 5) Many patients passed traditional tests of executive frontal lobe function but still failed the MET–SV. This pattern was strongly associated with observed dysexecutive symptoms in everyday life. The results demonstrate the clinical utility of the test, and suggest that there are two common and independent sources of failure on multitasking tests in a general neurological population: memory dysfunction, and initiation problems. (*JINS*, 2003, *9*, 31–44.)

Keywords: Ecological validity, Dysexecutive syndrome, Multitasking, Multiple errands

INTRODUCTION

It is ten years since Shallice and Burgess (1991) demonstrated that patients with frontal lobe damage may be impaired in everyday life situations requiring planning and multitasking despite normal or supernormal performance on traditional neuropsychological tests of language, memory, perception, and even executive functions. The explanation developed at that time and subsequently (e.g., Burgess, 2000; Burgess et al., 2000) was that the cardinal situations where these patients had problems are those that require subtle planning and prospective memory, and are "ill-structured" (Goel & Grafman, 1997) in the sense that there are many different ways to approach the task and participants have to decide for themselves how they allocate their efforts.

Shallice and Burgess (1991) described two procedures that were sensitive to their patients' particular dysexecutive problems. One was the Six Element Test, a measure of multitasking ability that can be conducted in the laboratory. This was subsequently developed and published as part of a battery of tests that measure executive functions (Wilson et al., 1996). We now know some of the basic clinical characteristics of this type of task (e.g., Burgess et al., 2000; see Burgess, 2000 for review), and a number of variants are starting to appear in the literature (e.g., Levine et al., 1998; Manly et al., 2002). There are two key sets of findings concerning this class of test. The first is that the frontal lesions that most consistently impair performance tend to be located in the most anterior and medial aspects of the frontal lobes (Burgess, 2000; Burgess et al., 2000; Levine et al., 1998; see also Burgess et al., 2001, for related functional imaging evidence). The second is that test performance is a good predictor of problems with planning and "intentionality" (i.e., goal-directed behavior) in everyday

life (e.g., Burgess et al., 1998), in other words that the test has good "ecological validity."

The second task used by Shallice and Burgess (1991) has received less attention. This was the Multiple Errands Test, a shopping task carried out in a pedestrian precinct. Participants were required to achieve a number of simple tasks without breaking a series of arbitrary rules. Shallice and Burgess's three patients were less efficient in the way they organized their efforts, completed fewer tasks and broke more rules than a group of healthy controls matched for age and intellectual ability. In addition performance of the brain-injured people was qualitatively unusual, and they made types of errors not seen in the controls.

This study demonstrated the MET to be potentially useful for the assessment of high level executive impairments in people whose general level of cognitive ability was "superior" and who passed existing frontal and other tests of cognitive functioning. However, its applicability to the range of people more typically encountered in routine clinical work remains unknown. Moreover, while the different forms of error made by patients on the test were considered, the relation between them, which would have informed our understanding of the demands of multitasking was not possible within a single-case design.

The first aim of the present study was therefore to produce a simplified version of the Multiple Errands Test (MET–SV) more suited to general clinical use, and determine its ecological validity. The second aim was to examine the relationship between different forms of error on the test to ascertain the minimum number of theoretically separable factors determining multitasking performance. We have tried throughout to implement the task in such a way that others should be able to adapt it for their own environment.

METHOD

Research Participants

Two groups of people participated in the study. The first consisted of 46 people who had no history of neurological disease. These were recruited from amongst members of staff employed by St Andrew's Hospital, Northampton, UK, and their associates. They were aged between 21 and 58 years (mean = 29.2, SD = 8.5). An indicator of general cognitive ability was obtained through performance on the National Adult Reading Test—Revised (NART–R: Nelson, 1991), which estimated the Wechsler Adult Intelligence Scale—Revised (WAIS–R: Wechsler, 1981). FSIQ for this group lay between 85 and 124 (mean = 107.6, SD = 9.1).

Equal numbers of male and female participants took part (n = 23). There were no significant gender differences regarding either age (t = 0.89, p = .379) or NART FSIQ (t = 1.35, p = .183).

Controls rated themselves using the self-completion version of the Dysexecutive Questionnaire (DEX-S: Burgess et al., 1996a). Ratings are made on a Likert-type scale of

twenty of the most commonly reported dysexecutive symptoms (e.g., disinhibition, confabulation).

Fifty people with brain injury also took part (41 male, 9 female). Thirty-six were inpatients admitted to a neurorehabilitation unit (Kemsley Division). The remaining cases were outpatients. Seventy-eight percent had sustained traumatic brain injury (TBI) as a result of motor vehicle accidents. Of the remainder, 9 had incurred brain damage through cerebrovascular accident and two had undergone surgery for cerebral tumors. Severity was determined with reference to duration post-traumatic amnesia, duration coma, or depth of coma when first admitted to hospital (see King, 1997). Most were categorized as "very severe" (75%). Of the rest, four were classified as "moderate," four as "severe," and one "mild." It was not possible to determine severity of injury for one participant. Mean time since injury was 72.1 months (SD = 68.4, minimum = 6, maximum = 372).

Age ranged from 18 to 59 years (mean = 34.6, SD = 12.7). Estimates of optimal pre-morbid levels of general cognitive ability were obtained using the NART–R, and indicated a mean FSIQ equivalent of 99.9 (SD = 12.9). Current levels of ability were somewhat lower (mean WAIS–R FSIQ = 84.1, SD = 12.7).

Brain injured participants completed the self-completion version of the Dysexecutive Questionnaire (DEX-S; Burgess et al., 1996). Patients were also rated on the second version (DEX-O: Burgess et al., 1996a) of this questionnaire by someone who knew them well. For inpatients, the DEX-O was completed by a member of the clinical team. Outpatients were assessed by a family member or carer. No one was available to rate two outpatients. As part of their routine neuropsychological examination, patients also completed one or more of the following: 1) the Behavioral Assessment of the Dysexecutive Syndrome battery (BADS: Wilson et al., 1996); 2) Cognitive Estimates Test (Shallice & Evans, 1978); 3) a verbal fluency test (FAS: Benton, 1968); and 4) a modified version of the Wisconsin Card Sorting Test (MWCST: Nelson, 1976).

Test environment

The MET–SV was conducted at a shopping center in Northampton, typical of those found in urban areas in the UK. Only the upper floor of this two level facility was used. Shops mainly comprised nationwide retail outlets selling a wide range of goods. One side of the center was dominated by a large supermarket, and there was a library and post office. An elevated clock could be seen from most locations (see Figure 1). Written permission to carry out the study was obtained.

Description of the MET–SV

This comprised a simplified version of the procedure described by Shallice and Burgess (1991) incorporating three principal modifications. The first was the provision of more concrete rules to enhance task clarity and reduce the likeli-
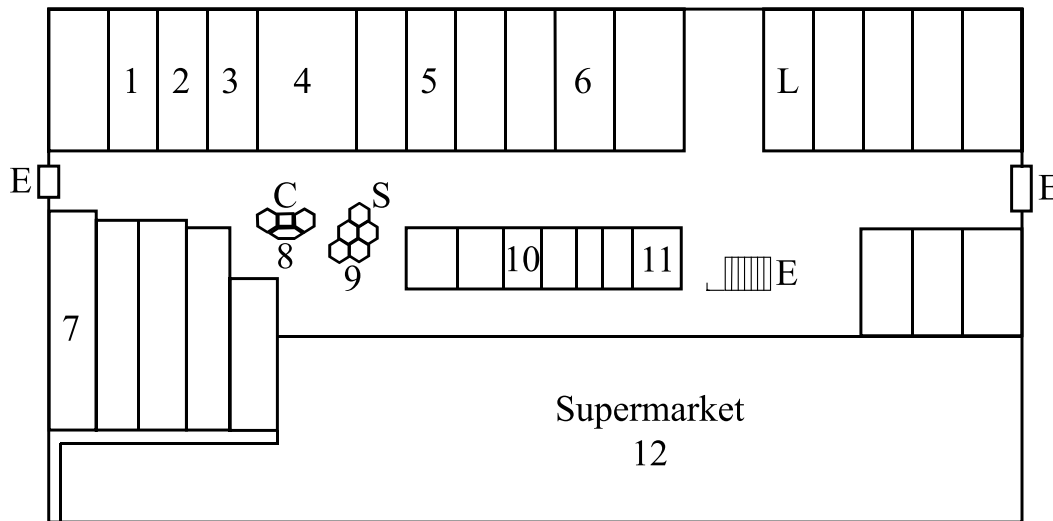
Fig. 1. Plan view of the upper level of the shopping center. Key: enclosed areas = shops; 'E' = exits; 'C' = clock; 'S' = stalls; 'L' = library. Numbers refer to locations where tasks can be accomplished. Items to be purchased: bread [12]; plasters [6, 7, 12]; birthday card [1, 3, 4, 5, 7, 12]; chocolate [4, 6, 7, 8, 12]; light bulb [1, 7, 12]; key ring [1, 2, 3, 5, 9, 10, 11, 7, 12]. Information to be recorded: newspaper headline [4, L, 12]; library closing time [L]; price of tomatoes [12]; shops selling TVs [12]. Meet under clock after 20 minutes [C]. Inform examiner when completed exercise [anywhere].

hood of interpretation failures. The second was simplification of task demands; and the third was the instruction sheet given to participants had space into which the participant was informed to record the information they were required to collect. In the Shallice and Burgess study, participants were not prevented from recording information, but this strategy was not suggested to them. (See Appendix 2.)

a) tasks

There were four sets of simple tasks analogous to those in the original MET, totaling twelve separate subtasks. The first involved purchasing six items (e.g., a light bulb; see Appendix 2). The second involved locating and recording four items of information (e.g., the closing time of the library on Saturday; see Appendix 2). The third was meeting the assessor at a designated point (under the clock) and stating the time 20 minutes after beginning the test. The last task was for the participant to tell the assessor when they had finished.

b) rules

To reduce ambiguity and simplify task demands, the number of rules was expanded from six in the Shallice and Burgess 1991 version to nine; these were also made more explicit and were clearly presented on the instruction sheet (see Appendix 2).

Procedure

Participants were accompanied to the shopping center by one or more assessors. Before starting, two ratings were obtained. First, participants were asked to rate the statement "How efficient would you say you were with tasks like shopping?" using a ten-point Likert type scale with weighted end points ("0"—"hopeless," "10"—"excellent"). Next, a measure of familiarity with the shopping center was made using a four-point scale ("0"—"never visited," "1"—"visited once or twice," "2"—"visit occasionally," "3"—"visit regularly").

The test began under the clock. Participants were given the instructions on a clipboard, a pen, carrier bag, ten pound (sterling) note, and (if necessary) a wrist watch. The instructions were read to the participants and the rules explained with reference to the instruction sheet. The boundaries of the shopping center were demonstrated clearly (demarcated by doors at either end). It was emphasized that the assessor would be following the participant at a distance to observe performance and that they should not be spoken to unless this was a specified requirement of the exercise. Next, the participants were prompted to ask any questions they had and then asked to explain what they were required to do. If necessary, instructions were repeated until the assessor was satisfied the participant knew what was expected. Finally, the start of the test was signaled by the instruction "Begin the exercise." (See Appendix 1.)

The assessor shadowed the participant and made notes but did not initiate interaction unless this was necessary on the grounds of safety or legality (e.g., if a participant attempted to leave the designated boundaries of the shopping center, or attempted to leave a shop without paying for items).

At the end of the test the participant was asked to rate the question "How well do you think you did with the shopping task?" using a ten-point scale with weighted end points ("0"—"hopeless," "10"—"excellent").

Analysis of individual participant performance

Participant performance was reviewed by at least two assessors (at least one of whom had been present during the test). Errors were categorized using the Shallice and Burgess (1991) definitions, that is: 1) *inefficiencies*—where a more effective strategy could have been applied; 2) *rule breaks*—where a specific rule (either social or explicitly mentioned in the task) was broken; 3) *interpretation failure*—where the requirements of a particular task were misunderstood; 4) *task failure*—a task not completed satisfactorily. Because the provision of an expanded set of rules made task demands less ambiguous, it was anticipated fewer inefficiencies and interpretation failures would be made and most errors would be straightforward to categorize. However, there were occasions when it was not obvious which category an error belonged to (e.g., when a patient went for a cup of coffee). These were resolved through discussion and the outcome recorded.

In addition, the number of times help was asked for (other than from the assessor conducting the test, as this comprised a rule breaking error) was also recorded and included in the analysis.

RESULTS

Performance of Control Participants

Error statistics for the control group are presented in Table 1. Floor effects were avoided (only one person did not make any errors at all). Most controls made seven or fewer errors (89.1%). However, five people made an excessive number of errors in comparison to the rest of those tested (9–16). The number made was disproportionate among this small group, accounting for 62/202 of the total errors recorded (30.7%). Of these, the majority comprised rule breaks (47) with just eleven percent of controls responsible for forty-two percent of all rule-breaking errors.

Effects attributable to age, sex, and NART–R in healthy controls

The number of category-specific errors made, and the total number recorded, were correlated with: age; NART–R FSIQ;

the number of times help was asked for; ratings of familiarity with the shopping center and with shopping in general; how well participants rated their performance; and the DEX total self-rating.

MET–SV errors were not significantly related to intellectual ability as estimated by the NART–R. However, there were positive correlations between age with rule breaks ($r = .31$) and the total number of errors made (.27). Interestingly, requests for help was also predictive of overall performance (.40), with more pleas associated with poorer performance.

Despite significant correlations between age and rule breaks, and rule breaks and requests for help (.41), older people did not seek help more (.17, ns). Females tended to accomplish fewer tasks than males, with the difference attaining borderline significance (male mean = 1.17 [SD = 0.71], female mean = 1.65 [SD = 0.89]: $t = 2.01$, $p = .050$). There were no significant gender differences regarding other errors or requests for help.

Seeking help was quite common, and was undertaken by over half the controls (54.3%): 57 appeals were made by 25 people. The majority asked for help once or twice (76%), although one person made nine requests (see Table 1). There was no difference between the mean number of task failures for those that sought help and those that did not ($t = 0.11$, ns). However, the mean number of rule breaks was significantly higher among people who asked for help (3.36, $SD = 3.20$) compared to those that did not (1.29, $SD = 2.19$: $t = 2.51$, $p = .016$). People who sought help also made more total errors (5.44, $SD = 3.65$) than those that did not (3.14, $SD = 2.63$: $t = 2.40$, $p = .021$).

Effects of familiarity with shopping environment in controls

Familiarity with the shopping center was investigated to determine if increased knowledge of the environment was associated with fewer overall errors. Participants were assigned to one of four familiarity categories, determined by their rating of how frequently they had visited previously. There was no significant discrepancy between the proportion of participants assigned to each of the four categories

Table 1. MET–SV performance amongst neurologically healthy ($N = 46$) and brain injured participants ($N = 50$)

	Neurologically healthy participants					Brain injured participants				
	Mean	*SD*	Minimum	Maximum	Total	Mean	*SD*	Minimum	Maximum	Total
Inefficiencies	0.39	0.61	0	3	18	0.58	0.84	0	4	29
Interpretation Failures	0.17	0.38	0	1	8	0.14	0.45	0	2	7
Rule Breaks	2.41	2.95	0	12	111	6.90	7.24	1	28	345
– actual	2.39	2.89	0	12	110	6.52	6.97	0	28	326
– social	0.02	0.15	0	1	1	0.38	1.10	0	7	19
Task Failures	1.41	0.83	0	3	65	5.46	2.70	1	12	273
Total Errors	4.39	3.40	0	16	202	13.08	7.78	3	35	654
Requests for Help	1.24	1.78	0	9	57	4.90	5.73	0	31	243

(chi-square = 1.83, ns). A series of one-way ANOVAs failed to show a significant effect of familiarity on the individual error categories, or the total number of errors made. Moreover, people less familiar with the shopping center did not ask for help more frequently. Similarly, there was no association between ratings of familiarity and participant's perceptions of their abilities with tasks like shopping, or indeed, between how well they thought they had completed the MET–SV and self-ratings on the DEX.

There were no gender differences regarding ratings of familiarity, efficiency with tasks like shopping, or perception of MET–SV performance. However, males believed they had more dysexecutive problems in everyday life (DEX-S male mean = 21.7 [SD = 8.1], female mean = 16.2 [SD = 9.5]: t = 2.03, p = .049). Curiously, this self-criticism was not justified by the results here, as males tended to make fewer errors than females.

Comparisons Between Brain-injured and Control Participants

Differences between group means for each error category were examined using a series of t-tests (see Table 1). Separate variance estimates were used for those comparisons where Levene's test indicated this was appropriate (see Howell, 1997, pp. 198–199). There were no significant differences regarding inefficiencies or interpretation failures. However, the mean frequency of errors made by brain injured participants was significantly higher than those achieved by controls for rule breaks (t = 4.03, p < .001), task failures (t = 10.10, p < .001) and total errors (t = 7.18, p < .001). (These results all survive Bonferroni correction for multiple comparisons.)

It is possible these differences might result from the small group differences in age, and familiarity with the shopping center noted between groups. This second variable is especially noteworthy as patients indicated they were less familiar with the environment (most indicated they had "never visited" before [74%] or "visited only once or twice" [24%]).

Group comparisons were therefore repeated using age, familiarity, and NART–R scores as covariates. Homogeneity of variance was achieved for each comparison where this was appropriate by transforming data using the square-root method (see Howell, 1997, p. 327). The previous results were preserved: (rule breaks, F = 11.79, p = .001; task failures, F = 40.82, p < .001, total errors, F = 27.92, p < .001).

Differences in patterns of error between patients and controls

As noted above, the groups made very different numbers of errors: a typical brain-injured participant made approximately three times the number of errors as an average control. However the relative proportion of error types were similar across groups. Rule breaks were most common

(55.0% and 52.8% of total errors for controls and patients, respectively). Task failures were the next most frequent (controls 32.2%, 41.7% patients). Rank order of proportions for the remaining error types (inefficiencies and interpretation failures) were also similar (see Table 1).

There was however one interesting difference in performance pattern between groups. While only one social rule break was observed within the control group, 19 were committed by the patients, which is statistically significant (chi-square 4.3, df = 1, p = .04).

Detailed analysis of error type

So far error analysis has followed the basic categorization provided by Shallice and Burgess (1991). The sensitivity of this method to neurological damage is demonstrated in the degree to which it can discriminate between the two groups. Using a cut-off of 12 or more errors (i.e., 5%ile of controls), 22/50 (44%) of brain-injured participants would have been correctly classified as patients, while 4.7% controls would have been misclassified. However, deriving a single measure based only on the number of errors fails to take into account qualitative differences in performance between groups. Shallice and Burgess (1991) remarked their three patients performed both quantitatively *and qualitatively* worse than their controls (p. 735) although this was not examined directly. Supporting this view, in this study 29 individual error types (as contrasted with categories) were demonstrated by controls (see Table 2), but almost double this number were observed within the patient group (58). Twenty-five error types were shared by both groups. Four additional types appeared only in the controls performance. Critically however, brain injured participants demonstrated 33 error types unique to them (see Table 3). For example, 54% failed to record the library closing time (a task failure), while no control did this. Others were less numerous: for example, 4% purchased items not on the list (interpretation failure). More surprisingly, one patient bought a pornographic magazine which, needless to say, is not on the list of items participants are requested to buy, and is classed as a social rule break.

This suggested a more sensitive scoring method could be devised. Accordingly, one was formulated that reflected the "normality" of the error type similar to the method used by Shallice and Evans (1978). "Normal" or acceptable errors were those that had been seen in up to 95 percent of healthy controls. These were given a score of "1." Errors only demonstrated by five percent or less of controls were assigned a weighted score of two. Errors unique to the patient group were given a weighted score of three.

Error scores were recalculated and a total weighted error score obtained. As before, a cut-off score derived using the 5th percentile of controls was used. This proved to be the same, 12 or more: however, 82% of the brain injured group were now correctly categorized compared to 4.7% of the controls being misclassified. The controls mean total score was 4.76 (SD = 3.71), while that for patients was 21.70

Table 2. Errors demonstrated by neurologically healthy participants (Note: any errors made by participants with acquired neurological disorders not shown here were assigned a weighted score of 3)

	Total participants who made error	Percent	Weighted error score
Task Failures			
Failed to inform examiners had finished test	13	28.3	1
Failed to meet examiner under clock	12	26.1	1
Met examiner under clock too early (<19 minutes)	11	23.9	1
Incorrect number of shops selling TVs recorded	11	23.9	1
Failed to inform examiner of time when under clock at correct time	6	13.0	1
Purchased non-birthday card	3	6.5	1
Failed to write down price of tomatoes	2	4.3	2
Purchased white loaf	2	4.3	2
Met examiners under clock too late (>21 minutes)	1	2.2	2
Purchased cream instead of plasters	1	2.2	2
Failed to purchase chocolate bar	1	2.2	2
Failed to write down newspaper headline	1	2.2	2
Failed to purchase bread	1	2.2	2
Inefficiencies			
Entered library (opening/closing times displayed outside)	11	23.9	1
Purchased multiple single items (e.g., 2 lightbulbs)	4	8.7	1
Purchased tomatoes to determine price per pound/kilo	1	2.2	2
Purchased medium brown loaf	1	2.2	2
Went to cafe for coffee	1	2.2	2
Rule Breaks-actual			
Entered shop and left without making a purchase	20	43.5	1
Spoke to examiner	16	34.8	1
Spent over £5.00	10	21.7	1
Re-entered shop previously visited	1	2.2	2
Purchased items from the 'stalls'	1	2.2	2
Purchased more than two items in superstore	1	2.2	2
Posted own (private) letter	1	2.2	2
Rule Breaks-social			
Asked staff in superstore to bake bread	1	2.2	1
Inefficiencies			
Wrote down purchase code of birthday card	4	8.7	1
Believed necessary to complete entire test within 20 minutes	3	6.5	1
Wrote down price of tomatoes in both pounds and kilos	1	2.2	2

($SD = 11.51$), a highly significant difference ($t = 9.87$, $p < .0001$). Effects of weighting are shown in Table 4.

Effects of age, IQ and requests for help on patient performance

As with controls, the brain injured group showed evidence of age effects, with a modest correlation between the number of task failures and age ($.25$, $p = .040$). In addition, WAIS–R FSIQ also correlated with task failures ($-.32$, $p = .030$). Critically however, the difference between patients and controls in task failures remained after covarying for age, FSIQ and familiarity with the shopping center ($F = 11.57$, $p = .006$).

While in controls, requests for help were significantly associated with poorer performance as measured by number of rule-breaks, patients showed a different pattern, with a negative correlation between help requests and task failures ($-.32$; $p = .013$). We will return to this later. A higher proportion of patients used the strategy of asking for help (80% *vs.* 54.3% controls).

Ratings of own performance in patients and controls

Among controls there was no significant relationship between their ratings of everyday executive abilities (DEX-S score) and MET–SV performance. In contrast, there were

Table 3. Errors unique to brain injured participants

	Total participants who made error	Percent
Task Failures		
Failed to purchase lightbulb	19	38
Failed to purchase birthday card	9	18
Failed to purchase keyring	17	34
Failed to purchase plasters	9	18
Purchased large brown loaf	6	12
Failed to write down closing time of library	27	54
Failed to write down number of shops selling TVs	19	38
Failed to complete writing down all of newspaper headline	1	2
Incorrect closing time of library recorded	2	4
Incorrect price of tomatoes recorded	5	10
Inefficiencies		
Left clipboard in shop	1	2
Left bag in shop	1	2
Ate chocolate prior to finishing test	1	2
Wrote down both opening and closing times of library	1	2
Wrote down closing time and telephone number of library	1	2
Wrote down two newspaper headlines	1	2
Purchased newspaper to obtain headline	1	2
Purchased items bought in same shop separately	3	6
Obtained and used shopping trolley	1	2
Ran out of money necessitating examiner intervention	1	2
Rule Breaks-actual		
Left the designated boundaries of the shopping center	6	12
Rule Breaks-social		
Overtly interrupted shop staff interaction	2	4
Shouted question to shop staff from outside shop	1	2
Shouted question to shop staff	1	2
Jumped queue at payment point	1	2
Swore loudly	1	2
Removed shoes	1	2
Blocked library stairs	1	2
Bumped into another person	1	2
Threw clipboard	1	2
Purchased pornographic magazine	1	2
Overtly invaded shop staff's personal space	1	2
Interpretation Failures		
Purchased item not on list	2	4

several significant correlations between patients ratings and performance on the test. However, these demonstrated lack of awareness, rather than presence of it: low ratings on the DEX-S were associated with increased rule breaking behavior ($-.30$, $p = .022$), task failures ($-.25$, $p = .047$) and more total errors made ($-.37$, $p = .005$). In other words, those who were poorer at the test were those who saw themselves as having few dysexecutive problems in everyday life.

Ecological Validity of the MET–SV

The MET is inherently ecologically valid in that it is little more than a formalized version of an activity in which people naturally indulge (in Western society). However its clinical utility would be enhanced if it could be shown that performance on the test is representative of performance outside the situation of shopping. In order to examine this possibility we determined characteristic behavior patterns

Table 4. Unweighted and weighted error scores by category and group

	Neurologically healthy participants		Brain injured participants	
	Unweighted error score	Weighted error score	Unweighted error score	Weighted error score
Inefficiencies	0.39	0.48	0.58	1.12
	(0.61)	(0.89)	(0.84)	(2.06)
Interpretation Failures	0.17	0.20	0.14	0.28
	(0.38)	(0.45)	(0.45)	(1.20)
Rule Breaks	2.41	2.50	6.90	8.64
	(2.95)	(3.02)	(7.24)	(8.84)
– actual	2.39	2.46	6.52	7.38
	(2.89)	(2.91)	(6.97)	(7.74)
– social	0.02	0.04	0.38	1.26
	(0.15)	(0.29)	(1.10)	(3.49)
Task Failures	1.41	1.59	5.46	11.62
	(0.83)	(1.07)	(2.70)	(7.16)
Total Errors	4.39	4.76	13.08	21.70
	(3.40)	(3.71)	(7.78)	(11.51)

on this test by conducting an exploratory factor analysis of the different error categories using all participants. This suggested a two-factor structure which accounted for 62.1% of the variance (Varimax rotated). Participants' MET–SV performance was characterized by either rule breaking behavior *or* failure to achieve set tasks. The first factor consisted of rule breaks and interpretation failures, and accounted for 35.7% of the variance (eigenvalue = 1.426). The second factor contained task failures and inefficiencies: 26.4% of the variance was explained by this factor (eigenvalue = 1.056).

Accordingly, patients were divided into two subgroups according to those errors made. For each person the number of rule breaking errors and interpretation failures were added together to provide a single score (labeled "Pattern 1"), while a further index was calculated by tallying task failures and inefficiencies (Pattern 2). People whose Pattern 1 score was greatest were categorized together ($n = 19$) while those Pattern 2 score was highest formed a second group ($n = 26$). Five participants achieved equal Pattern 1 and Pattern 2 scores and were excluded from subsequent analyses.

Some 574 errors were recorded. The proportion of errors made was reasonably similar between groups: thus, those

people who showed predominantly Pattern 1 accounted for 309 errors (53.8%) while those whose Pattern 2 score was greatest made a total of 265 (46.2%) errors. Rule breaks accounted for 72.5% of all errors made within the Pattern 1 group ("Rule Breakers"), and 74.2% of all rule breaking errors made by participants with acquired neurological damage were made by members of this group. By contrast, performance of the 26 people with high Pattern 2 scores was characterized by task failures which accounted for 64.5% of all errors made within the group, and 72.5% of all errors of this type (Table 5 shows a breakdown of error types for the two groups). The total mean error score for "Rule Breakers" ($n = 19$) was 16.26 ($SD = 10.96$) while that for the "Task Failers" ($n = 26$) was 10.19 ($SD = 3.78$), a difference that proved significant ($t = 2.31$, $p = .031$, separate variance estimates used).

We then considered the severity of the two groups' everyday dysexecutive problems as assessed by carers and relatives using the DEX questionnaire (Burgess et al., 1996a, 1998). Burgess et al. (1998) suggest that the DEX measures five principal symptom clusters: Inhibition, Intentionality, Executive memory, Positive affect, and Negative affect. Accordingly, we compared the severity of the five symptom clusters between the two groups (Rule Breakers and Task Failers). For two of the symptom clusters, the difference proved significant. Rule Breakers showed more severe *executive memory* symptoms, whose principal items are confabulation, deficits in temporal sequencing (i.e., mistaking the order of events) and perseveration ($t = 2.13$, $p = .04$). By contrast, Task Failers showed more symptoms of "negative affect" (e.g., shallow affect and apathy; $t = 2.30$, $p = .027$).

Requests for help amongst "rule breakers" and "task failers"

The influence of group (neurologically healthy *vs.* acquired neurological damage) and error style ("rule breakers" *vs.* "task failers") on requests for help were investigated. All participants who could be categorized by error style were included in the analysis ("rule breakers": controls $N = 20$, brain injured $N = 19$ [mean requests for help 2.20 *vs.* 7.77]; "task failers": controls $N = 23$, brain injured $N = 26$ [mean requests for help 1.00 *vs.* 2.86]) using a 2×2 ANCOVA in which familiarity was entered as a covariate. Both main effects were significant (group: $F = 11.55$, $p = .001$; error style: $F = 12.61$, $p = .001$). Interaction effects attained borderline significance ($F = 3.82$, $p = .054$).

Table 5. Distribution of MET–SV error types amongst the two subgroups of brain injured participants

	Inefficiencies	Interpretation failures	Rule breaks	Task failures	Total errors
Group 1: Rule Breakers	14	6	224	65	309
Group 2: Task Failers	15	1	78	171	265

The observation that brain injured people frequently ask others for help while carrying out the MET–SV appears to reflect an exaggeration of a strategy noted amongst controls, especially for "rule breakers." But does it help performance?

Correlations were computed for the two patient subgroups determined by error style. These suggested the effect of help seeking behavior on MET–SV performance differed. Asking for help was perhaps an aid to those participants whose performance was characterized by achievement of less tasks: those who sought assistance attained more tasks than those who did not (correlation between requests for help and weighted task failures = −.42, *p* = .017). However, while asking for help does appear to benefit MET–SV performance among "task failers," it appears to be of no benefit to people whose performance is characterized by rule breaking (correlation between requests for help and weighted rule breaks = .72, *p* < .001, and total weighted errors = .48, *p* = .023). One interpretation of this result is that the subgroup of participants who most employ help seeking as a strategy are those who derive the least benefit from it.

Relation between MET–SV and other executive function tests in patients

The distinction between two error styles was supported by the differential pattern of Person correlations with other executive test scores. Given the number of correlations involved we will consider only those that were significantly different from zero at the *p* < .01 level. Across all patients, number of task failures was related to MWCST perseverative errors (*r* = .39) and BADS profile score (−.46). The latter chiefly reflected individual correlations with two subtests: Zoo-Map (−.46) and Six Element Test (−.41). By contrast, rule breaks of both types (i.e., task rule breaks and social rule breaks) were related only to performance on the Action Program subtest of the BADS (−.40 social; −.42 task; −.43 total rule-breaks) although there was also a marginally significant relationship between Cognitive Estimates and one of the rule-break measures (social rule-break, −.33, *p* < .05).

In order to determine the sensitivity of the MET-SV in comparison to other executive function tests, the following procedure was adopted:

First, the proportion of patients who passed each of the three traditional experimentally-derived tests (Cognitive Estimates, Verbal Fluency, MWCST) was determined. Cut-off scores (below the 5% level of controls) were identified from data reported in two previous studies (Wilson et al., 1996; Burgess et al., 1998). The results were surprising. Despite the neurological group consisting mainly of people whose damage was classed as "severe," most patients who were given them passed these tests (Cognitive Estimates 93.3%; FAS Verbal Fluency 80.9%; MWCST categories achieved 89.5%, total errors 84.2%, perseverative errors 92.1%).

Twenty-nine patients completed all three frontal lobe tests: Seventeen (61%) achieved scores within acceptable limits on all of them. However on the MET–SV, only 5 patients achieved error scores that lay within the normal range. This suggests that the MET–SV is especially sensitive to executive function difficulties. A counter-argument might be, however, MET–SV failures can be caused by non-executive cognitive deficits to which the traditional executive tests are less sensitive. We can test this possibility. If it were true, we would not expect MET–SV failures to be strongly associated (over the group as a whole), with everyday observed dysexecutive symptoms (since the possibility of strongly correlated executive and non-executive deficits can be excluded in this group—see results of partialling for FSIQ above). Consequently, the various MET–SV error categories were correlated with the five factors plus the total sum of ratings from the DEX questionnaire.

Table 6 shows that MET–SV error category scores were highly predictive of ratings of executive symptoms in the group of patients who passed the traditional executive tests but failed the MET–SV shopping task. Significant correlations were found between errors and four DEX factors, and the total sum of ratings. Disinhibition and difficulties with formulating and following goal-oriented plans were especially associated with rule breaking behavior. Interestingly, ratings concerning positive affect (variable motivation, aggression, and euphoria) specifically related to social rule breaks, a finding which has good face validity. While WAIS–R FSIQ did not correlate with errors, age was associated with task failures. However, the pattern of associations remained unchanged when partial correlations were calculated in which variance attributed to age was controlled (unweighted task failures with *intentionality*, partial *r* = −.70; weighted task failures with *intentionality*, partial *r* = −.61).

DISCUSSION

An important goal of this study was to simplify the MET to render it more suitable to use with the wide range of people encountered in clinical practice. The performance of a group of neurologically healthy participants of predominantly average ability suggests this aim has been achieved. Floor effects have been avoided, and the test shows a remarkably high group difference between the patients and controls. Even using the simplest scoring method, the mean performance of the patient group was almost three standard deviations below the control mean. And using the most sensitive scoring method, 82% of the patients fell at or below the 5% performance level of the controls. This degree of sensitivity is remarkable, especially given that the actual test duration is only of the order of 45 minutes. Results suggest the MET–SV is considerably more sensitive for this purpose than many other executive tasks (see Burgess, 1997; Burgess et al., 1998). In this way, results continue to support the recent trend in neuropsychological assessment toward more "ecologically valid" procedures (e.g., Levine et al.,

Table 6. Correlations between error categories and the DEX-O for 17 brain injured participants who passed traditional executive function tests but failed the MET–SV ($p < .05$, 1-tailed)

	Factor 1 Inhibition	Factor 2 Intentionality	Factor 3 Executive Memory	Factor 4 Positive Affect	Factor 5 Negative Affect	DEX-O Sum of Ratings
Unweighted Errors Scores						
Inefficiencies		.59			.76	
Interpretation Failures	−.67					−.57
Total Rule Breaks	.70	.60				.64
Actual Rule Breaks	.66	.64				.59
Social Rule Breaks				.79	−.59	
Task Failures	−.58			−.52		
Total Errors		.67				
Weighted Error Scores						
Inefficiencies					.50	
Interpretation Failures	−.67				−.57	
Total Rule Breaks	.65	.52		.66		.69
Actual Rule Breaks	.68	.71				.66
Social Rule Breaks				.64		
Task Failures	−.57			−.52		

1998; Manly et al., 2002; Robertson et al., 1994; Wilson et al., 1985, 1989, 1997).

The results of this study also strongly indicate that even patients with severe brain injury may do well on traditional executive function tests, while still presenting with a dysexecutive syndrome as assessed both by a real-world shopping task and general behavioral observation in everyday life. This dissociation was first shown in patients who achieved FSIQs at or above the "superior" level (Shallice & Burgess, 1991). It is remarkable however to find a similar pattern in a group of patients whose injuries are more severe, and whose performance on cognitive tests is more variable. It is likely that the explanation for the frontal/ MET performance dissociation made by Shallice and Burgess also applies to patients of more mixed ability. While the patient group studied here were relatively heterogeneous, we would argue this is characteristic of many people referred for neuropsychological examination in routine clinical practice. Nevertheless, it would be useful to study if patients with mild and moderate severity of injury perform on the MET–SV in a similar way to the three "superior" level patients described by Shallice and Burgess and the more severe, variable sample reported here.

A further result of importance for the clinical utility of the test is that the MET–SV appears to have independence from the potentially biasing effects of general intellectual ability, familiarity with the environment and gender. None of these was associated with the performance of controls. With regard to patients, there was a modest correlation between the current level of intellectual ability (WAIS–R FSIQ) and task failures. However, the finding that patients make significantly more errors of this type than controls was preserved when variability ascribed to FSIQ was statistically controlled. This key result runs counter to the suggestion

that the MET–SV failures in our patients were principally due to non-executive problems. In any case, even if there had been an association, it is entirely plausible that the WAIS–R was being failed for executive reasons well tapped by the MET–SV and which also affect WAIS–R performance. In this context, the finding of group differences after covarying for FSIQ seems even more remarkable.

Equally interesting were the findings concerning two variables that were predictive of performance. The first, age, was a significant factor in the performance of both controls and patients. The second variable was requests for help; however, the nature of this relationship was unexpected. In controls more requests was associated with greater numbers of rule breaks, with no significant relationship with task failures. In patients, this *positive* association between requests for help and rule-breaks also appeared, but was stronger. However there was a significant *negative* correlation between requests for help and task failures. It is tempting therefore to see this result as suggesting a general performance principle: that people who readily ask for help are also those who tend to break task rules. This cannot be explained by suggesting that those asking for help are finding the task more difficult: they are actually achieving more tasks. However within the group who tended to fail to carry out tasks, those who asked for more help tended to do better. One possible explanation for these results is that asking for help is beneficial for both groups in terms of achieving set goals, but may also be an indicator of a general tendency towards disinhibition. Whatever the actual reason, it is clear that use of this strategy was not determined by age, general ability, gender, or environmental familiarity. It was also not associated with any of the three sets of ratings made.

The findings relating to these ratings of performance were also instructive. For the controls, ratings of their abilities

bore little relationship to their actual performance. Most believed they had performed satisfactorily, and at debriefing participants were often unaware of their errors. This would provide one explanation of the lack of the accuracy of their ratings. However, the pattern among patients was quite different in that those that performed poorly also tended to be those who rated themselves as having few executive problems in everyday life. This result is difficult to explain in terms of unawareness or error on the MET-SV. Instead it is suggestive of a more widespread problem with accurate assessment of executive abilities (i.e., "insight"). However this impairment is not one of applying a consistent standard, as one might expect if the root of the problem were an inability to use the scale, or engage in the judgement task itself. A problem of this kind would not result in a significant negative correlation. Instead, it suggests an overconfidence in abilities. But whether this overconfidence is a root cause, or is secondary to an inability to notice errors has yet to be established.

An anecdotal example may however help. PD, originally described by Burgess et al. (1996b), had incurred a severe brain injury as a result of a fall and showed a wide range of dysexecutive problems (see also Alderman, 1996). On the MET–SV his total weighted error score was 44; the majority of these were rule breaks (28). He also frequently asked for help (16). On one occasion during the test, he left the supermarket, stopped, and consulted his written instructions. He then looked up and appropriately sought assistance from a passer-by. He enquired as to the whereabouts of the library (one of the test items requires the participant to record the closing time of the library on Saturday). The passer-by told him and then added the helpful information ". . . I don't know if its still open—*I know it shuts at 4:30 on a Saturday* but I don't know when it closes during the week." PD offered his thanks and proceeded to the library where he spent some time finding this very same information. At debriefing he recalled asking for help in this instance, but was unaware that he had also been told the very information he was seeking. This example seems more consistent with a monitoring deficit than a problem of overconfidence.

This study also confirmed the proposition made originally by Shallice and Burgess (1991) that not all forms of error are equivalent in terms of their significance. While some errors are typical, others are characteristic of acquired neurological damage. For example, no neurologically healthy participant failed to write down the closing time of the library which is perceived as quite an important aspect of the test by controls: in contrast, this error was demonstrated by over half the patients (54%). Other task failures more characteristic of brain injury included the failure to purchase certain items (e.g., a lightbulb [38%]; failure to purchase a key ring [34%]; and failure to write down the number of shops selling televisions [38%]). However some patient errors were much more idiosyncratic. For example, one participant sat down and took off his shoes, while another threw his clipboard: no healthy participant engaged in similar behavior!

A key finding in this respect concerns social rule breaks, of which the two instances described above are examples. Although the patients made more than three times the number of general errors as the controls, they made nineteen times the number of social rules breaks. Even allowing for the relatively small numbers involved in this comparison, this difference was significant. These errors are therefore of particular diagnostic significance. The advantage of tasks such as the MET–SV is that they involve free interactions and social behavior that may be observed. This is not a feature typical of executive function tests and yet these results suggest this aspect of behavior is particularly instructive.

A further finding is that there appear to be two distinct patterns of MET–SV failure. People who show the first pattern tend to break task rules. These patients are those in whom carers tend to notice memory disturbances, especially those relating to memory control (e.g., confabulation, problems with temporal sequencing, and perseveration). Certainly a pure amnesia would not be a sufficient explanation of these people's difficulties since they have the task instructions with them at all times. Moreover, the amnesia explanation still does not work even if one suggests that this hypothetical basic memory problem would lead to forgetting to consult the instruction sheets. Such a performance would be characterized by incomplete tasks, and this was not the case. It appears not to be the case that they fail to consult the instructions, or that they forget what they want to do. Instead it appears to be that they fail to act upon the information they receive (as in the case of PD), or misunderstand what the instructions intend. This pattern has similarities with other failures in everyday life previously called "monitoring" errors (e.g., Alderman, 1996, 2001; Alderman et al., 1995).

The second pattern of impairment was characterized principally by task failures. Patients who showed this pattern were those in whom carers and relatives noticed negative affective symptoms, such as apathy and lack of emotion. An obvious explanation for this behavior is lack of motivation. However this group are conscientious about following the rules, which are generally quite effortful (e.g., having to avoid spending more than £5, considering carefully whether one needs to enter a shop before doing so), and terms of overall errors their performance was actually better than the "Rule Breakers." It seems likely therefore that a lack of emotional reactivity is a root cause, a deficit which is commonly linked with frontal lobe damage (e.g., Brazzelli et al., 1994; Stuss et al., 1992). However, since neuroscientific theories of the relation between motivation and emotional reactivity are only recently beginning to emerge (e.g., Posner & Rothbart, 1998; Rolls, 2000), this may be all that can be concluded at the present time. We will provisionally propose therefore that the root cause is a failure of "initiation," since at the most basic level the failures were most often quite simply that the tasks were not initiated. This is a relatively common symptom of frontal lobe damage (e.g., Burgess & Shallice, 1996).

The results clearly indicate that these two impairments ("monitoring" and "initiation") in this heterogeneous neurological group are orthogonal. Indeed, the quite extreme degree of categorization that could be achieved perhaps suggests that they are mutually exclusive, which may give a clue to the nature of them.

The distinction between the two patterns of MET–SV failure was supported by the differential relationship that these measures showed with scores from the other executive tasks. Thus, three individual test measures (MWCST perseverative errors; Zoo-Map test and the Six Element Test (SET)) were significantly related to number of MET–SV task failures. These results may inform a characterization of the underlying processing deficit. The Wisconsin Card Sorting Test has been argued to have a "working memory" component (e.g., Kimberg & Farah, 1993). The Zoo-Map test is a test of planning (participants are asked to plan a trip round a zoo), and the SET is a multitasking test with a strong prospective memory component as well as planning demands (Burgess, 2000). The MET–SV is also a multitasking test (see Burgess, 2000 for a description of the characteristics of this class of task). These results are therefore congruent with the model of multitasking presented by Burgess et al. (2000), where this function is supported by three cognitive constructs: memory, "intentionality" (largely prospective memory), and planning.

The relationship between the Action Program Test and rule-breaking behavior is less straightforwardly explained. Relatively little is known about the demands made by the Action Program Test. However it is assumed to have a judgement and reasoning component (Klosowska, 1976). This is also true of the Cognitive Estimates Test (Shallice & Evans, 1978). Both tests also involve initiating this unprompted consideration process before acting. This may be the link with rule-breaking behavior. However this speculation remains to be investigated.

A potential weakness of the present study is that the exact contribution of non-executive cognitive impairments other than those measured by the WAIS–R is not known. Another important point concerns the applicability of the method reported here to other settings, an important consideration to clinicians who want to utilize the test.

We have addressed both points in a further study which adapted the MET–SV for use in a typical hospital setting. Knight et al. (Knight, 1999; Knight et al., 2002) found a correlation between the two test versions of .83 ($p < .001$). This suggests that if clinicians were to apply the methodology of the MET–SV to their own patients within a local shopping facility, it is likely that broadly comparable results would be obtained. Knight et al. also administered a wider range of cognitive tests than given here, with the results suggesting that non-executive deficits (e.g., visual perception) were not a major factor in test performance. Only memory impairment, as measured by lower profile scores on the RBMT (Wilson, Cockburn & Baddeley, 1985) was also associated with a tendency to fail more set tasks, and recent functional imaging evidence in particular strongly

shows the involvement of executive frontal lobe processes in memory (e.g., Burgess et al., 2001).

The present study also does not address the issue of the brain localization of the processes relevant to shopping. However there is evidence from other studies that have used multitasking tests. Burgess et al. (2000) administered a variant of the Six Elements Test (Burgess & Shallice, 1991; Burgess et al., 1996) to a group of 60 acute neurological (mainly tumor) patients. This task required following a set of arbitrary rules and self-initiated task switching, and in this way overlaps with the MET–SV. Burgess and his colleagues found three sources of failure: memory problems, planning failures and failures of "intentionality" (i.e., following rules and attempting all the necessary tasks). Of course it is possible that the locus of these patients' impairments was not the same as with this (predominantly head-injured) group. However there is obvious similarity in the nature of the "intentionality" failures. Interestingly Burgess et al. (2000) did not distinguish between the two failures of omission and commission, with both found to result from lesions to the most anterior aspects of the frontal lobes (see also Levine et al., 1998). However the results of the present study suggest that these two patterns of failures might be considered as potentially separable.

These results perhaps suggest that different rehabilitation techniques may be useful. For instance, people who make task failures may be helped through training in the use of external aids or compensatory strategies, such as checklists (e.g., see Burke et al., 1991). Robertson, Levine, and their colleagues have also achieved promising results with people with everyday disorganization problems using Goal Management Training and other methods (Robertson, 1996; Levine et al., 2000; Manly et al., 2002). For the "Rule-Breakers," if our characterization of their impairments as problems with monitoring is correct, these have been shown to be resistant to some of the more frequently used neurobehavioral interventions (Alderman, 1996). However there are some indications that these people can be helped by methods that remediate or circumvent their monitoring problems (e.g., Alderman et al., 1995; Alderman & Knight, 1997).

ACKNOWLEDGMENTS

REFERENCES

Alderman, N. (1996). Central executive deficit and response to operant conditioning methods. *Neuropsychological Rehabilitation*, 6, 161–186.

Alderman, N. (2001). Management of challenging behaviour. In R.Ll. Wood and T. McMillan (Eds.), *Neurobehavioural disability and social handicap* (pp. 175–207). Hove, Psychology Press.

Alderman, N. & Knight, C. (1997). The effectiveness of DRL in the management and treatment of severe behaviour disorders following brain injury. *Brain Injury*, *11*, 79–101.

Alderman, N., Fry, R.K., & Youngson, H.A. (1995). Improvement of self-monitoring skills, reduction of behaviour disturbance and the dysexecutive syndrome: Comparison of response cost and a new programme of self-monitoring training. *Neuropsychological Rehabilitation*, *5*, 193–221.

Benton, A.L. (1968). Differential behaviour effects in frontal lobe disease. *Neuropsychologia*, *6*, 53–60.

Brazzelli, M., Colombo, N., Della Sala, S., & Spinnler, H. (1994). Spared and impaired cognitive abilities after bilateral frontal damage. *Cortex*, *30*, 27–51.

Burgess, P.W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81–116). Hove: Taylor and Francis.

Burgess, P.W. (2000). Strategy application disorder: the role of the frontal lobe in human multitasking research. *Psychological Research*, *63*, 279–288.

Burgess, P.W., Alderman, N., Wilson, B.A., Evans, J.J., & Emslie, H. (1996a). The Dysexecutive Questionnaire. In B.A. Wilson, N. Alderman, P.W. Burgess, H. Emslie, & J.J. Evans (1996). *Behavioural Assessment of the Dysexecutive Syndrome*. Bury St Edmunds: Thames Valley Test Company.

Burgess, P.W., Baxter, D., Rose, M.J., & Alderman, N. (1996b). Delusional Paramnesic Misidentification. In P.W. Halligan and J.C. Marshall (Eds.), *Method in madness: Case studies in cognitive neuropsychiatry*. Hove: Psychology Press.

Burgess, P.W., Alderman, N., Evans, J.J., Emslie, H., & Wilson, B.A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society*, *4*, 547–558.

Burgess, P.W. & Shallice, T. (1996). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, *34*, 263–273.

Burgess, P.W., Veitch, E.J., Costello, A. de L., & Shallice, T. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia*, *38*, 848–863.

Burgess, P.W., Quayle, A., & Frith, C.D. (2001). Brain regions involved in prospective memory as determined by positron emission tomography. *Neuropsychologia*, *39*, 545–555.

Burke, W.H., Zencius, A.H., Weslowski, M.D., & Doubleday, F. (1991). Improving executive function disorders in brain-injured clients. *Brain Injury*, *5*, 241–252.

Goel, V., Grafman, J., Tajik, J., Gana, S., & Danto, D. (1997). A study of the performance of patients with frontal lobe lesions in a financial planning task. *Brain*, *120*, 1805–1822.

Howell, D.C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Duxbury Press.

Kimberg, D.Y. & Farah, M.J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organised behavior. *Journal of Experimental Psychology: General*, *122*, 411–428.

King, N. (1997). Mild head injury: neuropathology, sequelae, measurement and recovery. *British Journal of Clinical Psychology*, *36*, 161–184.

Klosowska, D. (1976). Relation between ability to program actions and location of brain damage. *Polish Psychological Bulletin*, *7*, 245–255.

Knight, C. (1999). The development of a hospital version of the multiple errands test to assess executive functioning following severe acquired brain injury. Unpublished Doctoral Thesis. UK: The University of Birmingham.

Knight, C., Alderman, N., & Burgess, P.W. (2002). Development of a Simplified Version of the Multiple Errands Test for Use in Hospital Settings. *Neuropsychological Rehabilitation*, *12*, 231–255.

Levine, B., Robertson, I., Clare, L., Carter, G., Hong, J., Wilson, B.A., Duncan, J., & Stuss, D.T. (2000). Rehabilitation of executive functioning: An experimental-clinical validation of goal management training. *Journal of the International Neuropsychological Society*, *6*, 299–312.

Levine, B., Stuss, D.T., Milberg, W.P., Alexander, M.P., Schwartz, M., & Macdonald, R. (1998). The effects of focal and diffuse brain damage on strategy application: Evidence from focal lesions, traumatic brain injury, and normal ageing. *Journal of the International Neuropsychological Society*, *4*, 247–264.

Manly, T., Hawkins, K., Evans, J., & Robertson, I. H. (2002). Rehabilitation of executive function: Facilitation of effective goal management on complex tasks using periodic auditory alerts. *Neuropsychologia*, *40*, 271–281.

Nelson, H.E. (1991). *The National Adult Reading Test* (2nd ed.). Windsor, Berkshire: NFER Nelson.

Nelson, H.E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, *12*, 313–324.

Posner, M.I. & Rothbart, M.K. (1998). Attention, self-regulation and consciousness. *Philosophical Transactions of the Royal Society of London B*, *353*, 1915–1927.

Robertson, I.H. (1996). *Goal Management Training: A Clinical Manual*. Cambridge: PsyConsult.

Robertson, I.H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *Test of Everyday Attention*. Bury St Edmunds: Thames Valley Test Company.

Rolls, E.T. (2000). Memory systems in the brain. *Annual Review of Psychology*, *51*, 599–630.

Shallice, T. & Evans, M.E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex*, *14*, 294–303.

Shallice, T. & Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*, 727–741.

Stuss, D.T., Gow, C.A., & Hetherington, C.R. (1992). "No longer Gage": Frontal lobe dysfunction and emotional changes. *Journal of Consulting and Clinical Psychology*, *60*, 349–359.

Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale–Revised*. Texas, San Antonia: The Psychological Corporation.

Wilson, B.A., Cockburn, J., & Baddeley, A. (1985). *The Rivermead Behavioural Memory Test*. Bury St Edmunds: Thames Valley Test Company.

Wilson, B.A., Cockburn, J., & Halligan, P.W. (1989). *The Behavioural Inattention Test*. Bury St Edmunds: Thames Valley Test Company.

Wilson, B.A., Alderman, N., Burgess, P.W., Emslie, H., & Evans, J.J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome*. Bury St. Edmunds: Thames Valley Test Company.

Wilson, B.A., Evans, J.J., Alderman, N., Burgess, P.W., & Emslie, H. (1997). Behavioural assessment of the dysexecutive syndrome. P. Rabbitt (Ed.), *Methodology of frontal lobe and executive function* (pp. 239–250). Hove, UK: Psychology Press Ltd.

Appendix 1
MET–SV Task Instructions (read to participant)

MULTIPLE ERRANDS TEST

This test makes use of the following items:

- pen/pencil
- instructions on a clipboard for the subject
- carrier bag
- £10 note
- examiner to have stopwatch, pad and pen to record observations of subject

Ensure the subject is wearing a watch before leaving for the shopping centre.

Before leaving, obtain ratings for shopping efficiency and familiarity with the Weston Favell Shopping Centre (see separate sheet).

On return to the hospital, obtain the rating for how well the subject felt they executed the task (see separate sheet).

Once on the upper floor of the Weston Favell Shopping Centre proceed to the clock. Give the subject the clipboard, pen/pencil, carrier bag and £10 note. Read the following instructions to the subject:

"In this exercise I want you to complete three tasks. The tasks are: to buy the six items listed on this sheet (*examiner to indicate and describe items on sheet*); to obtain and write down four pieces of information (*examiner to indicate and describe items on sheet*); and to meet me under this clock 20 minutes after I have said " . . . begin the exercise" and tell me the time.

However, whilst 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 these tasks but you may do so in any order. You should spend no more than £5: although I've given you £10 you should spend no more than five. You should stay within the limits of the upper floor of the shopping centre. This means you must not leave by the doors at either end, use any lifts, stair or escalators. No shop should be entered other than to buy something, so if you go into a shop it should be with the intention of buying something. You should not go back into a shop you have already been in, so if you've been into a particular shop you should not go back into it again. You should not buy any items from the stalls: these are the stalls (show/gesture to subject) you must only buy items from shops. You should buy no more than two items from Tesco. Take as little time as possible to complete this exercise without rushing excessively.

During this 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 subject has*)

Now tell me what you must do. (*ensure subject is clear about what they must do*)

***"Begin the exercise*.**" (*examiner to start timing at this point*)

Appendix 2
MET–SV Written Instructions/Recording Sheet (given to and carried by participant)

INSTRUCTIONS

In this exercise you should complete the following three tasks:

1) You should buy the following items
small brown loaf	bar of chocolate
packet of plasters	single light bulb
birthday card	key ring

2) You should obtain the following information and write it down in the spaces below
   1. What is the headline from either today's 'Daily Mail', 'Daily Mirror' or 'The Sun' newspaper?
   2. What is the closing time of the library on Saturday?
   3. What is the price of 1 pound or kilogram of tomatoes?
   4. How many shops sell televisions?

3) You must meet me under the clock 20 minutes after you have started this 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 £5
► You should stay within the limits of the upper floor of the shopping centre
► No shop should be entered other than to buy something
► You should not go back into a shop you have already been in
► You should not buy any item from the stalls
► You should buy no more than 2 items in Tesco
► 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