# Ecological Assessment of the Supervisory Attentional System in People with Intellectual Disabilities

Tom Steverson

Doctoral Programme in Clinical Psychology

University of East Anglia

Word count: 32, 696

**Contents**

# List of Tables

## Abstract

The objectives of this study were to adapt a version of the Multiple Errands Test for people with intellectual disability and assess the degree to which it correlates with other measures of the Supervisory Attentional System, including the Tower of London Test, the Six Parts Test and the Dysexecutive Questionnaire. The study used a cross sectional correlational design with a single sample of 40 participants attending day centres for people with intellectual disabilities. The British Picture Vocabulary Scales – Third Edition, and the Word Reading subtest from the Wechsler Individual Attainment Test – Second Edition, were used to control for the potentially confounding effects of receptive vocabulary and reading ability on Multiple Errands Test performance. Results showed that ability to successfully complete tasks on the adapted Multiple Errands Test correlated significantly with the Tower of London Test. However, the adapted Multiple Errands Test failed to correlate significantly with the Six Parts test and the Dysexecutive Questionnaire. Also, performance on the adapted Multiple Errands Test was significantly related to receptive vocabulary, reading ability, and verbal IQ. The results suggest that the adapted Multiple Errands Test used here may expose impairment in the spontaneous schema generation, goal setting and adoption of processing mode (problem solving) functions of the Supervisory Attentional System in people with intellectual disabilities. This suggests that therapies aimed at remediating these deficits, such as Goal Management Training or Problem Solving Therapy, may help people with intellectual disabilities manage problems in executive function. However, it is concluded that the Multiple Errands Test used here needs further adaptation.

# Chapter One: Introduction

## 1.1 Overview

This introductory chapter presents the argument for the need to develop ecologically valid measures of executive function for people with intellectual disability (ID). Initially, a definition of executive function is given followed by a review of the relevant neuro-anatomical areas that are regarded as supporting executive functions. The functional and/or behavioural effects of deficits in executive function are then considered prior to an evaluation of three theoretical models. A critical discussion of traditional measures of executive function used to test these models is then presented followed by the argument that new measures need to be developed which can demonstrate ecological validity. The applicability of these models and measures for people with ID is then considered. An argument is proposed which supports the view that one particular measure may be a useful tool for assessing everyday functional and/or behavioural deficits observed in people with ID that could plausibly be attributed to deficits in executive function. Finally, the argument is reviewed and the aims of the study and research hypotheses are presented.

## 1.2 Defining Executive Function

Despite the frequency with which the term “executive function” is used within the neuropsychological literature, a unified definition is yet to be established (Jurado & Rosselli, 2007). For example, Spreen and Strauss (1998) described executive function as “a shorthand description of a multidimensional construct referring to a variety of loosely related higher order cognitive processes including initiation, planning, hypothesis generation, cognitive flexibility, decision making, regulation, judgement, feedback utilisation and self perception” (p. 171). Alternately, Burgess (2010) described executive function as “at the most basic level ... the abilities that enable a person to establish new behaviour patterns, and ways of thinking and to introspect upon them” (p. 349). At least at face value, these definitions appear to have only marginal common ground. This may, in part, stem from the debates ongoing within the executive function literature. For example, different perspectives exist concerning whether one underlying ability is responsible for performance on all executive tasks (e.g., Duncan, 1995), or whether executive functions are a set of fractionated abilities that can be impaired in isolation (e.g., Shallice & Burgess, 1996). This is also likely to be influenced by the way in which executive function has been measured, which has ranged from highly controlled and abstract, laboratory based experimental procedures (Demakis, 2003, 2004) to assessing people shopping in a supermarket (Alderman, Burgess, Knight, & Henman, 2003). These issues have led Jurado and Rosselli (2007) to conclude that there is currently no clear agreement on what executive functions actually are. However, for the purposes of this study, the following definition of executive function offered by Oscar-Berman and Marinković (2007) is adopted. This describes executive function as:

> "human qualities, including self awareness, that allow us to be independent individuals with purpose and foresight about what we do and how we behave. For example, executive abilities include judgement, problem solving, decision-making, and social conduct, and they allow us to monitor and change behaviour flexibly and in accord with internal goals and contextual demands" (p. 246).

This definition is employed as it combines both a descriptive account of some of the core cognitive processes that are regarded as falling under the banner of executive functions (e.g., self awareness, judgement, monitoring behaviour, problem solving and decision-making) as well as the behavioural functions that these cognitive processes serve (e.g., being independent, acting with purpose and foresight, social conduct, adapting to internal goals and contextual requirements). The neuro-anatomical regions that are regarded as supporting executive functions are now discussed.

## 1.3 Correlating Executive Functions with the Anatomy of the Brain

Much evidence has propelled the view that the frontal lobes are the regions of the brain in which the neuro-anatomical structures responsible for executive processes are located, leading many neuropsychologists to use the terms "frontal" and executive interchangeably (Baddeley, Della Sala, Gray, Papagno, & Spinnler, 1997). This evidence is largely based on the practice of measuring people with frontal lobe damage on tests that seemingly tap executive processes. For example, Demakis (2003, 2004) undertook two meta analyses comparing performance on four plausibly executive tests including the Wisconsin Card Sort Test (WCST; Heaton, Chelune, Talley, Kay, & Curtiss, 1993), the Categories Test (Benton, Hamsher, & Sivan, 1978), the Stroop Test (Stroop, 1935) and the Trail Making Test (Jarvis & Barth, 1984) in people with frontal and non-frontal lobe brain damage. Demakis found significant effect sizes differentiating poorer performance on the WCST, the Stroop Test and part A of the Trail Making Test in those with frontal lobe damage.

The weight of this evidence, and a wide body of corroborating research, clearly demonstrate that the frontal lobes have a vital role to play in executive functions (Stuss & Alexander, 2007; Stuss & Benson, 1984). Whilst accepting the value of the information gained from this approach, others have warned against the practice of correlating anatomical regions with cognitive function. For example, whilst Baddeley (1996) accepted that the frontal lobes are involved in executive functions, he outlines that other regions of the brain are also likely to play key roles in performance on executive tasks. Indeed, Baddeley et al. (1997) suggested that the distinction between frontal and non frontal tests is more a matter of degree than a true dichotomy, as it would be very difficult to design an experiment sensitive to only executive processes which do not tap areas such as language and perception. Moreover, whilst some studies have found consistently poor performance by people with frontal lobe lesions on executive tests (Burgess & Shallice, 1996a), impaired performance on the same tests can occur in those with brain damage outside of the frontal regions and even in neurologically healthy controls (Andrés & Van der Linden, 2001; Burgess & Alderman, 2004). Indeed, Burgess and Alderman (2004) described how a test score can only indicate a functional impairment that is associated with brain damage, not the biophysical state of a particular part of the brain. Furthermore, at least at the case study level, considerable damage to the frontal lobes as indicated by neuro-imaging can have a limited effect on neuro-behavioural test performance, whereas poor neuro-behavioural test performance can occur in the context of limited visible frontal lobe damage (Bigler, 2001). Moreover, Duncan, Emslie, Williams, Johnson, and Freer (1996) stated that caution must be exercised in assuming that poor test performance after brain damage is synchronous with the same functions that are measured when the test is undertaken by people from the normal population. Indeed, such tests may actually be measuring how humans perform on tests after brain damage, rather than impairment to the specific function that the damaged area purportedly performs.

In sum, the difficulty with the correlational approach is that correlation does not imply causation. That is, just because someone with frontal lobe damage performs poorly on a test of executive function, it does not necessarily mean that the two are causally related. Bearing this in mind, Baddeley et al. (1997) suggested that theoretical models of executive function offer a useful alternative to anatomical localisation as such models can be manipulated and tested experimentally rather than just sampled (as per the correlational approach). Several theoretical models of executive function are discussed in section 1.5. However, in order to help the models be interpreted and understood more adequately, it is useful to offer a prior explanation of how executive function impairment can manifest in an individual’s everyday presentation.

## 1.4 The Clinical Presentation of the Dysexecutive Syndrome

The behavioural presentation of people with deficits in executive function following frontal lobe damage has long been documented. For example, Rylander (1939) described such patients as experiencing "disturbed attention, increased distractibility, a difficulty in grasping the whole of a complicated set of affairs ... well able to work along old lines (but) cannot learn to master new types of tasks" (p. 22). Furthermore, Penfield and Evans (1935) described a patient who, after a frontal lobectomy, observed that:

> "She had planned to get a simple supper for one guest (WP) and four members of her family ... when the appointed hour arrived, the food was all there, one or two things were on the stove, but the salad was not ready, the meat had not been started and she was distressed and confused by her long continued effort alone ... she had become incapable of discerning for herself possible courses of action so that she might choose. If others presented the possibilities she made up her mind quite easily" (p. 131).

The cluster of functional problems that were experienced by the people in the examples above was traditionally defined as "frontal lobe syndrome". However, Baddeley and Wilson (1988) subsequently proposed that a more appropriate term might be "dysexecutive syndrome". This change in terminology represented a move away from the anatomical approach, to begin to describe the difficulties observed in people with frontal lobe damage in terms of functional impairment, as opposed to anatomical localisation. This was considered more useful to the practising clinician, who, with an understanding of impairment in functional or behavioural terms, is then much better placed to deliver interventions to help remediate the observed deficit(s).

In order to define the symptoms of the dysexecutive syndrome further, Burgess, Alderman, Wilson, Evans, and Emslie (1996) drew upon work by Stuss and Benson (1984) to develop the Dysexecutive Questionnaire (DEX). This samples four broad areas of likely problems that typically accompany frontal lobe impairment across the domains of emotion/personality, motivation, behaviour and cognition (Wilson, Evans, Emslie, Alderman, & Burgess, 1998). These are outlined in Table 1.1.

Table 1.1

*Characteristics of the dysexecutive syndrome as measured by the DEX questionnaire*

| Question | Characteristic assessed |
|---|---|
| 1. | Abstract thinking problems |
| 2. | Impulsivity |
| 3. | Confabulation |
| 4. | Planning problems |
| 5. | Euphoria |
| 6. | Temporal sequencing problems |
| 7. | Lack of insight and social awareness |
| 8. | Apathy and lack of drive |
| 9. | Disinhibition |
| 10. | Variable motivation |
| 11. | Shallowing of affective responses |
| 12. | Aggression |
| 13. | Lack of concern |
| 14. | Preservation |
| 15. | Restlessness-hyperkinesis |
| 16. | Inability to inhibit responses |
| 17. | Knowing-doing dissociation |
| 18. | Distractibility |
| 19. | Poor decision making ability |
| 20. | No concern for social rules |

*Note*. Adapted from “The Ecological Validity of Tests of Executive Function,” by P.W. Burgess, N. Alderman, J. Evans, H. Emslie and B. A. Wilson, 1998, *Journal of the International Neuropsychological Society, 4,* p. 549. Copyright 1998 by Cambridge University Press.

Whilst these features have been found in mixed aetiology neurological samples (Burgess, Alderman, Evans, Emsile, & Wilson, 1998), their presence in “normal” populations has also been observed (Chan, 2001). This suggests that the dysexecutive syndrome may be a matter of degree rather than a pathological entity, and that theoretical models of executive function may have much to offer the practising clinician in terms of understanding, formulating and providing interventions for a range of different clinical presentations which may present with seemingly dysexecutive symptoms. These models are now considered.

## 1.5 Theoretical Models of Executive Function

Whilst, a review of every model of executive function is beyond the scope of this project, three models were selected here on the basis of their relevance to the studies objectives. The interested reader is advised to consult Burgess and Robertson (2002) and Chan, Shum, Toulopolou, and Chen (2008) for more information about the different models of executive function.

### 1.5.1 Duncan's goal neglect theory.

Duncan's goal neglect theory (Duncan, 1986, 1995; Duncan, Burgess, & Emslie, 1995; Duncan et al., 1996; Duncan et al., 2008; Duncan, Johnson, Swales, & Freer, 1997) is based upon the importance of goal setting in human behaviour. Here, Duncan suggested that all human behaviours are goal-directed. Achieving these goals is dependent upon the ability to create smaller lists of goals or sub-goals. Therefore, when an individual's circumstances are experienced as sub-optimal (e.g., when experiencing hunger), a set of actions are internally developed or consulted to help bring the individual closer to their desired end state (e.g., relieving hunger by making a sandwich). However, actions incompatible with goal attainment can sometimes be activated by seemingly irrelevant or competing events (e.g., an important letter arrives whilst the sandwich is being made). The use of sub-goals therefore helps to impose some structure upon how the main goals are achieved and assists in inhibiting competing goals and activating relevant ones (e.g., eating the sandwich first and then reading the letter).

Duncan et al. (1996) further outlined how the commission of goal directed behaviour is largely regulated by the frontal lobes and suggested that many of the difficulties associated with frontal lobe damage can be accounted for by a single process referred to as goal neglect, or the "disregard of a task requirement even though it has been understood and remembered" (p. 257, Duncan et al., 1996). Duncan et al. (1996) also suggested that goal neglect and frontal lobe function(s) are strongly related to an individual's level of "*g*" or "general intelligence" (Spearman, 1927). Duncan et al. (1995) defined *g* as "a person's overall tendency to perform a task well or less well" (p. 262), and suggested that when a battery of diverse ability tests is administered to a large enough sample of people, a matrix of generally positive correlations will occur. Accordingly, Duncan et al. (1996) suggested that an underlying *g* factor is "the (hypothetical) factor responsible for broad positive correlations" on neuropsychological tests (p. 258) and, indeed, measures of executive function.

Duncan et al. (1995, 1996) stated that this contradicts the conventional view that executive functions are not related to psychometric "intelligence" (e.g., Eslinger & Damasio, 1985; Shallice & Burgess, 1991; Teuber, 1972; Warrington, James, & Maciejewski, 1985). Rather, Duncan et al. suggested that this view has come about due to problems in the way *g* has been specified and measured. Accordingly, they suggest that the concept of *g* is better related to that of fluid intelligence. Fluid intelligence is implicated in novel problem solving and is generally regarded as being assessed by tests such as Raven's Progressive Matrices (Raven, Court, & Raven, 1998) or Cattell's Culture Fair test (Institute for Personality and Ability Testing, 1973). Alternately, crystalised intelligence is regarded as knowledge that can be learned and is relatively insensitive to change. This is measured by tests such the National Adult Reading Test (Nelson, 1982), which is typically used to assess predicted premorbid IQ in people with dementia or brain injuries.

In support of their claims, Duncan et al. (1995) presented data from three patients who had cognitive deficits following frontal lobe lesions, but scored in the superior range on IQ tests. They measured these participants on Cattell's Culture Fair test to assess fluid intelligence. Two participants were also assessed on the seven sub test short form of the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) and one participant on the Wechsler Adult Intelligence Scale – Revised (WAIS-R; Wechsler, 1981) to measure crystallised intelligence. They found that whilst the participants with frontal lobe lesions had superior WAIS/WAIS-R IQ scores, comparable to those of neurologically healthy matched controls, the Cattell's Culture Fair test scores of those with frontal lobe lesions were 23-60 IQ points lower. Moreover, the discrepancies between WAIS/WAIS-R IQ and Culture Fair IQ were absent in five participants with brain lesions outside of the frontal lobe. Accordingly, because the participants with frontal lobe damage scored much more poorly on Cattell's Culture Fair test, they suggested that fluid intelligence or *g* is located in the frontal lobes.

Whilst the above study provides tentative evidence, the sample is small and the use of two different tests to measure crystalised intelligence (e.g., WAIS and WAIS-R) complicates the comparisons. It also makes no reference to goal neglect. Accordingly, Duncan et al. (1996) went onto investigate this further and relate the findings to goal neglect by using a letter monitoring task in which healthy participants were required to watch a particular side of a computer screen and read out the letters (but not the numbers) that appeared on that side. Occasionally, a prompt appeared asking the participants to switch sides. Of 90 neurologically healthy participants, 15 neglected this prompt (or as the authors suggested displayed goal neglect), despite understanding and remembering the prompt's requirements. These 15 people all scored one standard deviation below the mean on Cattell's Culture Fair test (a measure of fluid intelligence). This observation was replicated in a subsequent experiment using a sample of 41 elderly participants, aged between 60 and 70, whose mean Cattell's Culture Fair test score was biased towards a lower level of *g*. Here, low scores on Cattell's Culture Fair test were associated with poor performance on the letter monitoring (goal neglect) task ($r$ = .52). In contrast, correlations between the letter monitoring (goal neglect) task and two measures of crystallised intelligence; the Mill Hill Vocabulary Scale (Raven et al., 1988) and the Nelson-Denny reading comprehension test (Brown, Nelson, & Denny, 1976), were lower ($r$ = .02 and .30 respectively) suggesting that goal neglect was not related to crystallised intelligence in this sample but more closely linked with fluid intelligence. They also repeated these measures in 10 participants with frontal lobe lesions and 8 participants with posterior lesions, each with matched controls. They found no significant difference between fluid intelligence in those with posterior lesions and controls ($p$ =

0.8) but a significant difference between fluid intelligence in those with frontal lobe lesions and controls, with the participants with frontal lobe lesions showing worse performance ($p < .005$). Equally, on the letter monitoring (goal neglect) task there was a significant difference between participants with frontal lobe lesions and matched controls ($p < .005$) with complete neglect being shown by 7/10 participants with frontal lobe lesions and only one control. In contrast, there was an identical mean score between the posterior group and matched controls. Participants with frontal lobe lesions also scored significantly worse on the letter monitoring (goal neglect) task than participants with posterior lesions ($p < .02$). From this Duncan et al. (1996) suggested that both *g* and goal neglect were located in the frontal lobes.

Whilst these studies aim to provide some support for the theory that fluid intelligence or *g* is related to the function of the frontal lobes, across both studies only 14 participants with frontal lobe lesions actually took part. These small samples therefore make it difficult to generalise firm conclusions about the specific role of the frontal lobes in plausibly executive tasks, or fluid intelligence. Also, the study did not use any alternate tests of executive function so inferences about how much *g* or fluid intelligence actually contributes to executive function are difficult to draw. Therefore, Duncan et al. (1997) investigated these issues in a much wider sample of 90 brain injured participants. Here, they administered four tests traditionally considered to measure executive function. These included the Modified Wisconsin Card Sorting Test MWCST (Nelson, 1976), a verbal fluency test (Benton, 1968), a list learning task (Luria, 1966) and a spatial puzzle (Luria, 1966). They also administered five measures designed to tap areas unrelated to executive function. These included tests of Recognition Memory (Warrington, 1984), Digit Span (Wechsler & Stone, 1945), Object Naming (McKenna & Warrington, 1983), Recognition of Objects from Unusual Views (Warrington & Taylor, 1983) and Motor Speed (Annett, Hudson, & Turner, 1974). In the brain injured participants, correlational analysis showed that there were generally positive relationships between performance on the executive tests (between $r$ = .17 and $r$ = .34). However, the median correlation between these measures was small ($r$ = .26) and was very similar to the median correlation between the executive tests and tests of non executive ability ($r$ = .29). From this they suggested that such traditional executive tests were not strong measures of any stable underlying cognitive characteristic. Therefore, in a second study, Duncan et al. (1997) asked 24 of the above participants to complete a measure of fluid intelligence (Cattell's Culture Fair Test), the letter sequencing (goal neglect) test described above (Duncan et al., 1996), five plausibly executive tests, including a Verbal Fluency Test (Benton, 1968), Self Ordered Memory Test (Petrides & Milner, 1982), the Six Element test (Burgess et al. 1996), Tone Counting (Wilkins, Shallice, & McCarthy, 1987) and Reversal of Ambiguous Drawings (Ricci & Blundo, 1990) and two non executive tests including Recognition Memory (Warrington, 1984) and Digit Span (Wechsler, 1987). They found that the median correlation between the executive tests was low ($r$ = .04) whereas the median correlation between the executive and non executive tests was medium ($r$ = .38). Nevertheless, when correlations between all the tests were calculated, with fluid intelligence partilalled out, the correlation disappeared ($r$

= .04). Duncan et al. used this to suggest that little in common exists between these tests aside from their *g* or fluid intelligence component. Furthermore, a composite index derived from the executive tests correlated with fluid intelligence ($r$ = .59) and goal neglect ($r$ = .72). Duncan et al. therefore suggested that fluid intelligence and goal neglect are the factors underlying performance on executive tests.

Despite this, a correlation matrix between the severity of lesion in 24 brain injured participants across either frontal, temporal or parietal lobes (as rated on a scale of 0 - 4 by one of the researchers following examination of MRI scans), when correlated with the executive function, fluid intelligence and goal neglect scores, revealed generally low scores (ranging from $r$ = -.03 to $r$ = .36) and there is no mention of how many of the 24 brain injured participants had exclusively frontal, temporal or parietal lobe damage. These factors make conclusions about the specific relationship that *g* or goal neglect has to the frontal lobes (relative to other brain regions) tentative.

Equally, the measures used to assess goal neglect (Duncan et al. 1996, 1997) have been controlled experimental procedures that relate poorly to how goal attainment might be required in everyday life. For example, it is difficult to infer how reading letters from an alternate side of a computer screen might be related to formulating goals in the “real world”. This limits the extent to which inferences can be supported about how goal neglect phenomena might occur outside of the laboratory.

Furthermore, whilst it has been suggested that tests such as Raven’s Progressive Matrices or Cattell’s Culture Fair test are measures of fluid intelligence or novel problem solving ability (Ducan et al., 1995), Wilson, Evans, Emslie, Alderman, and Burgess (1998) suggest that “one of the critical aspects for measuring the dysexecutive syndrome is novelty” (p. 221). Thus, the skills required in good performance on novel tests such as those designed to tap executive function or fluid intelligence may not be entirely separate concepts. Indeed, Blair (2006) proposed that fluid intelligence, working memory and executive function form a unitary construct called fluid cognition and Kane and Engle (2002) have proposed “WM capacity, or the capability for executive attention, as the psychological core of the statistical construct of general fluid intelligence” (p. 638). Thus, some might consider Raven’s Progressive Matrices or Cattell’s Culture Fair test measures of executive function themselves, rather than fluid intelligence.

Despite these criticisms, the theory of goal neglect has stimulated research focused on the straightforward rehabilitative implications that it offers for goal setting in real life. Accordingly, Levine et al. (2000) used Robertson’s (1996) Goal Management Training (GMT; a rehabilitation method based on Duncan’s goal neglect theory) designed to assist people who present with difficulties in developing appropriate goals and sub-goals to learn the steps required to achieve set tasks. Levine et al. randomised 30 participants with brain injuries into either GMT or motor skills training (MST; training in processes unrelated to goal management) and assessed them on a series of “everyday” pencil and paper tasks (including marking words when proofreading, grouping items on a checklist and answering questions by making reference to a grid where the answers could be deciphered) both before and after the training. The GMT group showed significantly improved accuracy and took more time over the tasks than the MST group. They also described a case study in which GMT was found to be successful at improving the meal preparation skills of a participant with encephalitis. Whilst this is promising, there was a lack of evidence to suggest that GMT generalised to skills outside of the laboratory and no follow up was carried out to assess maintenance. Moreover, no matched control was employed in the single case study. Indeed, it is likely that this type of intervention would only be applicable to a specific sub group of brain damaged people motivated to complete the training and whose executive impairment was predominantly related to goal neglect.

Nevertheless, Burgess and Robertson (2002) outline a number of difficulties that underlie theories (such as Duncan’s goal neglect) which try to explain executive function as deriving from one single underlying factor. Initially, they suggest that, in group studies of either neurological or healthy samples, the correlations between executive tests are typically very low (Miyake et al. 2000; Robbins, 1998). Whilst Duncan’s argument is based on this finding (e.g., Duncan et al., 1997) an alternate perspective is that if one single factor were responsible for performance on all executive tasks (such as *g* or goal neglect), then these correlations might be expected to be much higher. Moreover, as executive function is such a multifaceted construct, different measures may tap different aspects of executive function respectively. For example, Verbal Fluency tests are regarded as measuring verbal production and the WCST is regarded as assessing switching or perseveration (Chan et al., 2008). With such a diversity of functions considered under the construct of executive function it is perhaps unsurprising that the correlations are low. Indeed, Miyake et al. (2000) used structural equation modelling to confirm that three often postulated executive functions including inhibition, shifting and updating contribute differentially to performance on separate executive tasks. Later, Friedman et al. (2006) also found that different aspects of intelligence correlate with these different aspects of executive function respectively. For example, in comparing measures of fluid intelligence, crystallised intelligence and the Wechsler Adult Scale of Intelligence – Third Edition (WAIS-III; Wechsler, 1998) with measures of inhibiting, shifting and updating in a neurologically healthy sample, they found that updating was highly correlated with all the intelligence measures, whereas inhibiting and shifting were not. Thus, there may be a much more complex relationship between the different aspects of executive function and intelligence than that proposed by Duncan.

Equally, Stuss and Alexander (2007) suggest that, rather than there being one underlying single factor contributing to general performance on executive tests, there are alternately three separable and distinct frontal attentional processes. Here, they outlined data from both neuropsychological test performance and functional imaging in those with frontal lobe lesions to identify energization, task setting and monitoring. Specifically, energization is proposed to be located in the superior medial region of the frontal lobes, task setting is positioned in the left lateral area and monitoring is associated with the right lateral region. This suggests a broader range of frontal lobe functions may exist as opposed to just one (e.g., goal neglect).

A further problem with Duncan's theory of goal neglect (albeit with the previous paragraph in mind) is the supposition that goal neglect and *g* are specifically located in the frontal lobes. As outlined above, poor performance on measures of executive function could occur for a range of reasons that are not necessarily related to frontal lobe function (Baddeley et al. 1997). These issues are equally compounded by Baddeley's (1996) suggestion that the idea of all executive functions being accounted for through a single underlying factor replaces one problem with another. That is, it is not clear what the true nature of *g* itself is, as most measures of intelligence sample a range of different skills which themselves are likely to tap a number of processes. Indeed, Duncan et al., (1996) suggest that the concept of *g* is itself, poorly specified or "abstract" (p. 258).

Overall, therefore whilst the strengths of Duncan's theory of goal neglect may be its parsimony, clearly the different abilities tapped by different executive measures and the range of anatomical correlates that they might plausibly relate to, make this theory open to criticism. This therefore paves the way for alternative theories of executive function, which have attempted to define executive function in more fractionated terms.

### 1.5.2 Baddeley's central executive.

The concept of the central executive derived from investigations into working memory. Working memory has been defined as "the system for the temporary maintenance and manipulation of information" (Baddeley et al., 1997) and has been perhaps most significantly influenced by Baddeley and Hitch's (1974) working memory model (see also Baddeley, 1986, 2001). Baddeley (2001) proposed that working memory comprises four mechanisms: a phonological loop which stores auditory information, a visuo-spatial sketchpad which stores visual information, an episodic buffer which acts as a general storage system that combines several different kinds of information and a central executive which resembles attention and deals with cognitively demanding tasks. It is the central executive that is most relevant to the field of executive function. Baddeley et al. (1997) suggested that the central executive coordinates the operation of its slave systems (the phonological loop and visuo-spatial sketchpad), acting as a general attentional resource involved in reasoning, decision making, calculation and long and short term information retention. Baddeley (1996) outlined the four functions of the central executive as (a) switching the retrieval of plans, (b) timesharing on dual task activities, (c) selectively attending to certain stimuli whilst ignoring others and (d) temporality holding and activating information from long term memory.

The central executive has perhaps been most extensively tested using dual task methodology (Baddeley et al., 1997; Della Sala, Foley, Beschin, Allerhand, & Logie, 2010). This follows the premise that if the central executive is responsible for commissioning the function of the phonological loop and visuo-spatial sketchpad, performance of two concurrent tasks that tap these slave systems simultaneously should result in impaired task performance (relative to when the tasks are performed individually) due to increased load on the attentional resources of the central executive.

Accordingly, based on research that people with Alzheimer’s disease demonstrate a unique profile of performance on verbal and visuo-spatial memory span tasks that differed from people with amnesia or head injury

(Spinnler, Della Sala, Bandera, & Baddeley, 1988; Wilson, Baker, Fox, & Kaszniak, 1983), Baddeley's working memory model was used as a framework to formulate this neuropsychological profile. In an early study, Baddeley, Logie, Bressi, Della Sala, and Spinnler (1986) asked 28 participants with Alzheimer's disease, 28 age and education matched controls and 20 "young" controls to undertake a computer based "pursuit tracking" task which was assumed to depend primarily on the visuo-spatial sketchpad. Once performance on the tracking task was established, participants were then asked to perform the original tracking task again, whilst simultaneously performing a secondary task. There were three secondary tasks including, (a) an articulatory suppression task where the participant was asked to count from one to five repeatedly at a rate of twice per second, presumed as relying primarily on the articulatory loop, (b) a reaction time to a tone task, where participants pushed a foot pedal in response to an auditory tone which was thought to place demands on attentional capacity (e.g., the resources of the central executive) and (c) a digit span task where the participant was asked to repeat different series of digits that increased in length, which purportedly made demands on both the central executive and the articulatory loop. Baddeley et al., hypothesised that performance on the original tracking task would become impaired if it were accompanied by the simultaneous performance of a secondary task and that this impairment would be greater for those with Alzheimer's disease compared to both age and education matched and young controls.

Analysis of variance for the primary tracking task showed that there was no main effect of group (e.g., Alzheimer's disease sufferers, elderly matched controls and young controls) on the primary tracking task performance, when performed without a secondary task, suggesting that all groups had comparable scores on this measure, $F < 1$. However, when the primary tracking task was combined with the secondary articulatory suppression task, there was a difference in tracking task scores across groups, $F(1, 54) = 4.71$, $p < 0.05$. Here, the young controls scored better than the Alzheimer's group, who actually scored slightly better than the elderly controls. However, there was no interaction effect between group (e.g., Alzhiermer's disease sufferers, elderly controls and young controls) and condition (e.g., tracking alone and tracking plus articulatory suppression), $F < 1$, suggesting that, for this condition, the decrement in tracking performance (when combined with articulatory suppression) was not dependent on whether the participants had Alzheimer's disease or not.

For the reaction time to tone condition, analysis of variance, as before, replicated the non significant differences between primary tracking task performances, when performed without a concurrent secondary task, across the groups, $F < 1$. However, when the primary tracking task was combined with the reaction time to tones task, there was a significant main effect of group (e.g., Alzhiermer's disease sufferers, elderly controls and young controls), $F(1, 54) = 23.53$, $p < .001$. There was also a significant interaction between group and condition (e.g., tracking alone and tracking plus reaction time to tones), $F(1, 54) = 6.13$, $p < .05$. This suggests that each group's tracking performance was affected differently by the addition of the secondary reaction time to tones task. Here, the participants with Alzheimer's disease showed worse tracking performance than the elderly controls, and both these groups showed worse tracking performance than the young controls. The interaction effect indicated that the decrement in dual task performance (relative to single task performance) was significantly greater for the participants with Alzheimer's disease, compared to the decrement in dual task performance (relative to single task performance) seen in the elderly and young controls. That is, the decrement in dual task performance (relative to single task performance) in participants with Alzheimer's disease was significantly steeper than the decrement in elderly and young controls performance, hence the crossover interaction.

For the digit span condition, analysis of variance, again, replicated the non significant differences between primary tracking task performances, when performed without a concurrent secondary task, across the groups, $F < 1$. However, when the tracking task was combined with the digit span task, there was a significant main effect of group (e.g., Alzheimer's disease sufferers, elderly controls and young controls), $F(1. 54) = 61.47$, $p < .001$, and a significant interaction between group and condition (e.g., tracking alone and tracking plus digit span), $F(1, 54) = 18.71$, $p < .001$. This again suggests that each groups tracking performance was affected differently by the addition of the secondary digit span task. Here, the participants with Alzheimer's disease performed much worse than the elderly controls, who themselves performed slightly better than the young controls. The interaction effect indicated that the decrement in dual task performance (relative to single task performance) was significantly greater for the participants with Alzheimer's disease, compared to the decrement in dual task performance (relative to single task performance) seen in the elderly and young controls.

That is, the decrement in dual task performance (relative to single task performance) in participants with Alzheimer's disease was significantly steeper than the decrement in elderly and young controls performance, hence the crossover interaction.

Overall, the results suggested that the combination of having Alzheimer's disease and undertaking a secondary task that loads onto the central executive (e.g., reaction time to tones and digit span), causes a particularly detrimental effect on primary task tracking performance. Accordingly, the Alzheimer's disease sufferers appeared to demonstrate a specific central executive deficit leading to difficulties coordinating two activities at the same time that was independent from the effects of normal ageing.

In order to test this further, Baddeley, Bressi, Della Sala, Logie, and Spinnler (1991) hypothesised that, if a central executive deficit does exist in Alzheimer's disease, retesting Alzheimer's participants on subsequent occasions would show that central executive function worsens as Alzheimer's disease progresses. Accordingly, using the same tracking, articulatory suppression, reaction time to tones and digit span tasks, they retested 15 of the participants with Alzheimer's disease, as well as 18 of the elderly controls that took part in the Baddeley et al. (1986) study, at both six and 12 month intervals. They hypothesised that there would be a difference between single and dual task performance and that this effect would share an interaction with (be dependent upon) the three testing sessions (e.g., scores from the original study, and scores at six and 12 months). Accordingly, in participants with Alzheimer's disease, there was no main effect of single task tracking performance across the testing sessions, showing that performance on the primary tracking task did not worsen over time. However, mean tracking performance in participants with Alzheimer's disease, when accompanied by the simultaneous performance of a secondary task, progressively worsened over the three time intervals in each condition and these main effects were significant for the articulatory suppression ($F(1, 14) = 39.5$, $p < .001$) and reaction time to tones ($F(1, 14) = 81.64$, $p < .001$) tasks. There were also significant interactions between the primary tracking task performance, when combined with a secondary task, across the suppression ($F(2, 28) = 4.4$, $p < .025$), reaction time to tones ($F(2, 28) = 18.57$, $p < .001$) and digit span ($F(2, 28) = 8.05$, $p < .01$) conditions. This indicates that, in participants with Alzheimer's, the decline in tracking performance was affected differently at different stages of the disease. That is, whilst single task performance did not progressively get worse over the three time intervals as Alzheimer's disease progressed, dual task performance (and plausibly central executive function) did progressively get worse over the three time intervals. When the same analysis was conducted on the elderly controls there was no worsening of scores over time or interaction effects. These findings provide further evidence of a specific central executive deficit unique to Alzheimer's disease sufferers that progressively becomes more impaired over time. This was later supported by Greene, Hodges and Baddeley (1995) who, using a dual task measure similar to that used by Baddeley et al. (1986) as well as a dual task measure taken from the Test of Everyday Attention (TEA; Robertson, Ward, Ridgeway & Nimmo-Smith, 1994) found that under both single and dual-task conditions, impaired performance occurred in those with Alzheimer’s disease compared to controls, with worse performance occurring in those with more advanced stages of the condition.

However, a difficulty with the interpretation of these findings is that any increase in task difficulty may result in impaired performance in those with Alzheimer’s disease as opposed to any specific central executive deficit under dual task conditions. To test this Baddeley et al. (1991) examined the effects of deterioration over time on a word categorisation task, where 30 Alzheimer’s participants had to match words to particular categories (e.g., animals, jobs, colours). Here, as the task progressed, the number of available categories increased in frequency, thus increasing the difficulty. It was hypothesised that, as such semantic categories do not appear to be heavily dependent upon working memory capacity any increase in the difficulty of a task should differentially impair Alzheimer’s patients and become more discernible as the disease progresses. Alternately, if dual task coordination is a unique deficit, there will be no interaction between increased difficulty and test session on the word categorisation task. A two way analysis of variance with test session and number of categories as repeated measures showed that whilst increasing the number of categories led to decreased performance in Alzheimer’s sufferers, the magnitude of the effect of task difficulty did not increase as the disease progressed as there were no interaction effects between the number of categories and occasion of testing. Accordingly, Baddeley et al. (1991) used this as evidence to suggest that there exists a specific central executive deficit in Alzheimer’s disease sufferers that is not related to increasing task difficulty.

Central executive impairments have also been observed in more ecological scenarios. For example, Alberoni, Baddeley, Della Sala, Logie, & Spinnler (1992) found that the ability to concentrate on more than one conversation was impaired in Alzheimer's disease sufferers compared to matched controls. Equally, Verghese et al. (2007) found that in elderly participants, the quality of walking gait reduced when they were asked to combine walking with a verbal letter sequencing task.

Aside from central executive deficits being well documented in Alzheimer's disease (see Huntley & Howard, 2009 for a review), studies using dual task methodology function have also revealed specific impairments in people with brain injury (Foley, Cantagallo, Della Sala, & Logie, 2010; McDowell, Whyte, & D'esposito, 1997) which has had implications for the rehabilitation of behavioural difficulties following brain injury. Here, Alderman (1996) assessed dual task performance in 10 neurologically healthy controls, 10 brain injured participants who responded well to a behavioural rehabilitation programme based on operant conditioning principles and 10 brain injured participants who responded poorly to the programme. It was found that whilst the brain injured participants were comparable on a range of neuropsychological measures, the "poor responders" showed consistently significantly worse performance on the dual task measures compared to "responders" and controls. Alderman explained this as a central executive deficit in the poor responders that may lead to behavioural problems and poor response to operant conditioning methods due to a failure to attend to simultaneous monitoring of their own behaviour and the contingencies of any reinforcement programme. Nevertheless, Alderman found that poor responders' performance on the primary task (a visuo-spatial tracking task) was improved by immediate verbal feedback about their performance from the experimenter. Alderman therefore suggested that immediate verbal feedback of behaviour would support the ability of the poor responders' central executive to focus on aspects of the environment that that they found difficult to attend to and therefore aid their learning ability and capacity to engage in the rehabilitation programme.

Moreover, aside from Alzheimer's disease and brain injury, dual task deficits have also been observed in people with HIV (Hinkin, Castellon, & Hardy, 2000), Schizophenia (Bressi, Miele, Bressi, & Astori, 1996; Junghoon, Glahn, Nuechterlein, & Cannon, 2004), ADHD (Kofler, Rapport, Bolden, Sarver, & Raiker, 2009), Parkinson's disease (Dalrymple-Alford, Kalders, Jones, & Watson, 1994; Malapani, Pillon, Dubois, & Agid, 1994), depression (Nebes et al., 2001), people with ID (Danielson, Henry, Rönnberg, & Nilsson, 2010) and people with Down's Syndrome (Kittler, Krinsky-McHale, Devenny, & Conners, 2008; Lanfranchi, Baddeley, Gathercole, & Vianello, 2011) suggesting specific central executive deficits in these populations. These findings provide further support for the validity of the central executive as a construct that can be weakened in the presence of injury, illness or disability.

However, despite its evidence and useful clinical implications, the majority of support for the central executive comes from studies employing the dual task paradigm methodology, with limited alternative experiments being used. Accordingly, a number of criticisms have been levelled at the methodology behind dual task studies. Firstly, Hegarty, Shah, and Miyake (2000) drew upon central bottleneck theory (Welford, 1952) which suggests that dual task slowing may actually be a consequence of the inability to perform certain mental operations simultaneously. For example, Pashler (1994) found that a bottleneck phenomenon can occur in tasks that require rapid response selection (e.g., where two rapid responses cannot be delivered simultaneously, thus creating a bottleneck effect leading to impaired performance). Accordingly, Hegarty et al. suggested that executive tasks that require rapid response selection such as rapid constant number selection (Baddeley, 1966), could interfere with a secondary task that involves the same processes via a bottleneck effect, as opposed to any specific central executive overload. Secondly, Hegarty et al. cite research by Bourke, Duncan, and Nimmo-Smith (1996) which found that in dual tasks, participants tend to allocate more resources to those tasks which they perceive as more demanding, regardless of which task is designated as primary. Thus, there may be a strategic trade off related to the complexity of the tasks. Therefore, it was suggested that, when applied to the central executive, dual task logic might not always apply.

Accordingly, using 120 neurologically healthy participants, Hegarty et al. (2000) developed three different primary visuospatial tasks that were thought to increase in difficulty as well as the demands they placed on the central executive. These included (a) a complex paper folding task that made the most demands on the central executive but required no rapid response selection, (b) a card rotating task that made a moderate demand on the central executive and required moderate response selection, and (c) a simple picture matching task that made the least demand on the central executive but required rapid response selection. They matched these primary tasks with two secondary tasks that were thought to load heavily onto the central executive and required rapid response selection. The two executive tasks included random number generation (vocalising random numbers between zero and nine) and a two back task where participants had to listen to a series of letters and say “yes” when they heard a letter that was identical to the letter presented exactly two items previously and “no” if it were not. They hypothesised that, according to standard dual task logic, concurrent performance of a secondary executive task should impair the primary paper folding task the most, as this itself was regarded as making the most demands on the central executive and would therefore cause a central executive overload. In contrast, if a response selection bottleneck or a strategic trade off was occurring then performance on the primary picture matching task (which made least demands on the central executive) would be most influenced by a secondary executive task that required rapid response selection, as the picture matching task required rapid response selection itself therefore causing a bottleneck effect. In addition, the picture matching task was relatively simple and consequently a strategic trade off might occur where the participant would allocate more attentional resources to the executive task, which was more complex, at the expense of the picture matching task, thus again impairing picture matching task performance.

Indeed, a one-way, within-subjects analysis of variance found a significant main effect for the difference between performance on the three primary visuo-spatial tasks when simultaneously performed with the secondary executive tasks of random number generation, $F(2,46) = 12.62$, $p < .01$, and the two back test, $F(2,46) = 12.23$, $p < .01$. Post hoc analysis showed that for both secondary tasks, the degree of detriment on the primary picture matching task was significantly greater than on the card rotations and paper folding tests (both $p$ <.05), which did not significantly differ from one and other. The authors suggested that these results supported the response selection bottleneck or strategic trade off hypothesis, where the addition of a secondary executive task that required raid response selection caused the greatest detriment on the picture matching task performance (hypothesised as loading least onto the central executive but requiring rapid response selection and posing a potential strategic trade off) and the least detriment on the paper folding task (suggested as loading most heavily onto the central executive thus as posing a dual task effect when accompanied by another executive task). From this they suggest that caution be exercised when interpreting the results of dual task studies and where the use of response selection or strategic tradeoffs can occur, researchers make efforts to control for these.

Nevertheless, it could also be argued that one reason why the executive tasks interfered with the picture matching task was because response selection is an important executive function governed by the central executive, and as the picture matching task required response selection, it was tapping an executive component itself (as well as a visuo-spatial component) therefore causing a dual task effect. The same argument could be applied for allocating attentional resources to a potential strategic trade off, which, it could be argued, might also be considered a function of the central executive. A further related difficulty with the study is objectively establishing the specific degree of central executive and/or visuo-spatial sketchpad involvement needed for successful performance on each of the three primary visuo-spatial tasks. One potential means would be to see how each of the measures used in the study correlate with other measures of the central executive or visuo-spatial sketchpad. Whilst the paper provides some evidence for the executive component in the visuo-spatial tasks (by citing correlations between the tasks and other measures of executive function), there is no mention of correlations between the tasks and other tasks of visuo-spatial ability. Therefore, the degree to which the visuo-spatial tasks were tapping the visuo-spatial sketchpad, above and beyond functions that could be accounted for by the central executive is unclear. Thus, as there is evidence that each visuo-spatial task has a central executive component, then this could support the dual task explanation of the results if the central executive demands involved in the task outweigh the visuo-spatial demands.

Salthouse et al. (1995) outline further methodological problems with dual task studies. For example, across many studies, the types of tasks used in combination have varied. This is concerning because different types of tasks may tap different “pools” of attentional resources and therefore, the extent to which each component task places equal demand on the central executive is not clear. This makes it difficult to meaningfully compare results across studies. Equally, they cite various studies which have focused on the performance of the primary task, and then examined the performance decrement on that primary task when combined with a secondary task, without measuring independent performance on the secondary task. This gives a misleading interpretation because some participants may find performing the secondary task much harder/easier than others and, without factoring this into the overall performance analysis, it can potentially distort primary task performance. One way around this has been to compute single task performance as a ratio of secondary task performance, however in the studies that have taken this approach, the mathematics used to compute the ratios have varied, thus further complicating comparison across studies.

Aside from this, criticisms have also sprung up in the literature attacking the central executive for being vaguely specified. For example, the central executive has been described as a “homunculus” (Allport, 1993) or “a little man who sits in the head and in some mysterious way makes the important decisions” (Baddeley, 1996, p. 6). Indeed, Parkin (1998) outlined that the central executive is often used to explain away functions that cannot be attributed to the phonological loop or visuo-spatial sketchpad, making it “a concept that emerges from research by default when more rigorous theoretical constructs cannot handle the data” (p. 519). In response to such criticisms, Baddeley (1986) adopted the supervisory attentional system (SAS; Norman and Shallice, 1986; Shallice and Burgess, 1996) as a more detailed specification of the specific operation of the central executive. This model is described below.

### 1.5.3 Norman and Shallice’s supervisory attentional system.

The SAS offers perhaps the most detailed specification of the multiple processes considered under the title of executive functions. At its simplest level, the Norman and Shallice model comprises two main components: the contention scheduling system and the supervisory attentional system. The contention scheduling system (CSS) governs the practice of routine behaviours by using incoming perceptual information to select the most appropriate schema (or organised plan) to suit a situation. These schemas are typically well rehearsed organised plans that can be deployed in routine situations with little conscious effort (e.g., driving the familiar route to one's workplace). As these schemas generally run automatically, the CSS acts to select, prioritise and implement the correct schemas, based on environmental demands. However, in order to prevent an irrelevant schemas being activated in inappropriate circumstances or what Shallice (1982, p. 201) defines as a "capture error" (e.g., accidentally driving the familiar route to one's workplace on a weekend when one actually intended to go to the supermarket), the SAS intervenes to bias the activation of schemas so they are appropriate to the situation. Thus the SAS contains "the general programming or planning systems that can operate on schemas in every domain" (Shallice, 1982, p. 201), before finally monitoring and evaluating how effectively the schemas are operating. Plausible accounts of a malfunctioning SAS can be clearly observed in the aforementioned anecdotal accounts of patients with executive dysfunction such as Rylander's (1939) description of those who are "well able to work along old lines (but) cannot learn to master new types of tasks" (p. 22) and Penfield and Evans's (1935) report of a patient who "had become incapable of discerning for herself possible courses of action so that she might choose. If others presented the possibilities she made up her mind quite easily" (p. 131).

One of the first tests of the SAS was undertaken by Shallice (1982) who adapted the traditional Tower of Hanoi test, into the Tower of London test (TOLT). The TOLT includes a rectangular board supporting three equidistantly placed pegs of ascending heights. Three coloured discs (with holes through their diameter) can be placed onto the pegs. The tallest peg can accommodate three discs, the middle peg can accommodate two discs and the shortest peg can accommodate one disc. Both the examiner and participant have identical versions of the board. The test starts with the examiner placing three discs onto the participant's board in a specified starting position. Behind a screen, the examiner then arranges three discs onto the pegs of his/her board to represent a desired goal state. The examiner then reveals his/her arrangement and the participant is required to replicate the examiners arrangement whilst obeying a set of rules. These include not exceeding the specified number of moves to reach the goal state, moving only one disc at a time, placing no more than one disc onto the shortest peg and two discs onto the middle peg, and ensuring that each disc is always placed onto one of the three pegs throughout the task and not put onto the table (Rainville et al., 2002). As the task progresses so do the number of moves required to replicate the examiners arrangement, hence increasing the task's difficulty. In order to complete the task successfully, the participant is required to divide the task into smaller subgoals and progressively work through these accordingly (Shallice, 1982; Simon, 1975).

Shallice (1982) adopted the TOLT as it had been used to test the planning and problem solving units of computer programmes and, as the SAS is regarded as performing these functions itself, it is plausible that such a measure would also test the planning and problem solving abilities of humans. Equally, successful performance on the TOLT does not require the use of any special purpose subroutines, thus relying exclusively on the SAS and nullifying the need for the CSS. Accordingly, those with an intact SAS would perform much better on the TOLT compared to those without, where it would be expected that distractibility and perseveration would occur (Shallice, 1982). To test this, Shalilce assessed 61 participants with brain injuries and 20 healthy controls on the TOLT. Using computer tomography (CT) scans, the brain injured participants were divided into lesion site (posterior or frontal) and hemisphere (left or right). Results showed that the participants with left frontal lesions completed significantly fewer problems on the TOLT than those with right frontal lesions, posterior lesions and controls (unfortunately actual significance values are not reported in this paper).

A difficulty with this study however, is that the TOLT is a specific test of the SAS rather than both the SAS and the CSS. However, this was overcome by the subsequent development of the Hayling Sentence Completion Test (Burgess & Shallice, 1996a). This requires the examinee to initially complete a sentence read out by the examiner with a word that plausibly makes sense (part A). For example, the examiner would read out “the captain went down with the sinking” to which the examinee would most logically respond “ship” (Chan et al., 2008). In part B the examinee is required to complete the examiners sentence with a word that is completely unconnected. For example, responding with the word “cow” to the sentence “the captain went down with the sinking” (Chan et al., 2008). On both parts the examinee is timed and needs to respond as quickly as possible. Shallice and Burgess (1996) suggest that performance on part A simply requires use of the CSS to initiate a logical response. However, to succeed in part B the participant must use the SAS to develop a novel strategy (or temporary schema) in order to produce a word prior to the examiner finishing the sentence (in order to respond as quickly as possible) and monitor its effectiveness. To test this Burgess and Shallice measured 47 participants with frontal lesions, 27 participants with posterior lesions and 20 controls. Here, on part A (straightforward sentence completion), a one way analysis of variance showed significant differences in the speed of responding across groups ($p < .002$). Post hoc analysis showed that the participants in the frontal group were significantly slower than controls ($p < .01$) whereas none of the remaining comparisons was significant at the $p < .05$ level. For section B (irregular sentence completion), a one way analysis of variance also showed a significant difference in errors between groups ($p < .002$). Here, post hoc analysis showed that the participants in the frontal group committed more errors than the controls ($p < .05$) and the participants with posterior lesions ($p < .01$). Nevertheless, on part B there were no differences in response latencies across the groups. Indeed, critically, there was no correlation between performance on part A compared to performance on part B in participants with frontal lobe lesions ($r = .19$, *ns*) whereas there was a significant correlation between the two parts in controls ($r = .59$, $p < .01$) and participants with posterior lesions ($r = .39$, $p < .05$) suggesting that the two tasks maybe tapping two separable processes: part A, response initiation as governed by the CSS and part B, suppression of a habitual response as governed by the SAS. Specifically the lack of correlation in the participants with frontal lesions suggests that the SAS may have been differentially impaired in many of these participants, hence no correlation between the two measures. In contrast, the correlation between parts A and B in the participants with posterior lesions and controls suggests that many of these participants had a functional SAS and CSS. This was confirmed by findings that four participants with frontal lobe lesions and one with a posterior lesion scored within one standard deviation of the mean on part A, but less than three standard deviations of the mean on part B, whereas one participant with a frontal lesion showed the opposite pattern (poor performance on part A but good performance on part B). This again suggested that two different types of strategy responses were being tapped by performance on parts A and B (e.g., the CSS and the SAS respectively).

However, despite this evidence the SAS was still itself vaguely specified and subject to the same criticisms that were levelled at Baddeley's central executive (Shallice, 2002). Thus, in order to refine the workings of the SAS, Shallice and Burgess (1996) updated the model. This outlined the SAS as responsible for a number of processes carried out by different subsystems, which work together in a globally integrated function (Shallice & Burgess, 1996). Shallice and Burgess outline these processes as including working memory, monitoring, rejection of schema, spontaneous schema generation, adoption of processing mode (problem solving), goal setting, delayed intention marker realisation (remembering to do something in the future) and episodic memory retrieval (drawing upon memory from past experiences).

The application of these processes is highlighted in the Brixton Spatial Anticipation Test (Burgess & Shallice, 1996b). This involves presenting the examinee with a continuous array of 56 pages of 10 circles arranged in a 2 x

5 matrix. One of the circles is always shaded and moves around the array in a predictable pattern (or rule) across each consecutive page. For example, the shaded circle may start by moving in a chronological order (e.g., 1, 2, 3, 4, 5), but intermittently the rule will change (e.g., 1, 5, 1, 5, 1, 5). The examinee therefore has to detect that a change in rule has occurred and modify their responses accordingly. Burgess and Shallice (1996b) measured 40 participants with frontal lesions, 24 participants with lesions outside of the frontal lobes and 20 controls on the Brixton Spatial Anticipation Test. Three different types of error score were plausible. These included (a) all errors that could be attributed to perseveration with a rule when it was inappropriate to continue with its use, (b) all other errors that could be attributed to the use of some incorrect rule (e.g., misapplication of a previously active rule) and (c) the number of times a rule is changed when there is no need to do so. Results showed that the frontal groups produced more errors across all types than posteriors, however this was only significant for type b ($p < .05$) and c errors ($p < .005$). Nevertheless, error types a and b correlated significantly with one another ($r = .60$, $p < .01$), however error type c correlated with neither a ($r = .13$) nor b ($r = .13$) errors. Shallice and Burgess (1996) interpreted this as suggesting that there were two separable factors on the Brixton test, which was supported by structural equation modelling. Here, the first factor (responsible for the correlation between error types a and b) was associated with an inability to produce a new hypothesis, or the strategy generation process, which they proposed as involving the aforementioned working memory, monitoring, rejection of schema, and spontaneous schema generation functions of the SAS. In contrast, the second factor, responsible

for error type c related to a malfunction in the adoption of processing mode (problem solving) function of the SAS by being in temporary search of an appropriate new schema, instead of being able to implement and activate a new schema.

Impaired performance on these measures (the TOLT, Hayling test and Brixton test) has been demonstrated in cognitively impaired groups such as those with schizophrenia (Marczewski, Van der Linden, & Larøi, 2001), Parkinson’s disease (Bouquet, Bonnaud, & Gil, 2003), Alzheimer’s disease (Lange, Sahakian, Quinn, Marsden, & Robbins, 1995) and older (relative to younger) people (Andrés & Van der Linden, 2000), thus further attesting to the validity of the theory.

Indeed, the SAS provides a much more comprehensive model of the multiple processes that might be undertaken by a central executive/supervisory system, thus overcoming criticisms of such a concept being a homunculus. Moreover, the SAS accounts for findings where correlations between tests purported to measure frontal lobe function are low, as the different tests may be measuring different aspects of executive function, which may be impaired in isolation (Shallice & Burgess, 1996). Indeed, Burgess et al. (1998) have shown that there are a range of different factors underlying the behavioural symptoms of the dysexecutive syndrome, as sampled by the DEX, rather than just one (e.g., goal neglect). Moreover, Burgess et al. showed that these different symptoms were consistent with different patterns of performance on executive tests. For example, different measures of executive function loaded significantly onto the DEX factor scales of inhibition, intentionality and executive memory respectively. This suggests that influences beyond goal neglect or *g* may be influencing performance on these measures and that in this case, different areas of the SAS might explain these impairments respectively. Thus, a multiple process approach such as that of the SAS accommodates these data well, not least in that it accounts for findings where individuals with frontal lobe lesions, who appear to perform well on many neuropsychological assessments, show very specific dysexecutive symptoms in the absence of others, such as deficits in decision making (Eslinger & Damasio, 1985) multitasking (Goldstein, Bernard, Fenwick, Burgess & McNeil, 1993; Shallice & Burgess, 1991) or confabulation (Burgess & McNeil, 1999). Here, it can be suggested that specific functions of the SAS are impaired in isolation.

On top of this, there has been a strong push amongst researchers to develop measures of the SAS that are more predictive of the types of dysexecutive problems encountered in everyday life. Two of these include the Multiple Errands Test (MET) and the Six Element Test (SET; Shallice & Burgess, 1991). The MET was originally designed to be carried out in a shopping centre and test the planning, multitasking and prospective memory capabilities of the SAS but in a “real life” environment. The SET represents a diluted, office based analogue for the MET but was developed to measure a subset of the same cognitive processes as the MET and thus remain sensitive to everyday dysexecutive problems (Burgess, 2000; Wilson et al. 1998). These measures and their theoretical links to the SAS are described in more detail in the proceeding sections.

More recent support for the SAS has been gathered through the model being the basis upon which test batteries such as the Behavioural

Assessment of the Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996; Wilson et al., 1998) have been developed. The BADS is a battery of six subtests which are designed to be more reflective of the types of executive skills required in everyday life including planning, self directed organisation, temporal judgement, set shifting, inhibition of established responses and novel problem solving (Manchester, Priestley, & Jackson, 2004). The BADS is widely used in clinical practice and has a large body of normative data. However, perhaps most promising is the data supporting the test battery's ecological validity. Here, Wilson et al. (1998) found that BADS total profile score correlates well with relatives or carers impressions of how participants with brain injuries function as rated by the DEX questionnaire, $r = -.62$, $p < .001$, where higher BADS score means better executive function, whereas higher DEX score means worse executive function. Equally, Bennett, Ong and Ponsford (2005) found that DEX ratings made by a neuropsychologist and occupational therapist significantly correlated with the BADS total profile scores of brain injured participants, $r = -.37$, $p < .01$ and $r = -.39$, $p < .01$, respectively. Also, Norris & Tate (2000) found that for three subtests of the BADS that were able to significantly discriminate between brain injured participants and neurologically healthy controls (Action Program test, Zoo Map test and the SET), these measures were also able to significantly predict Role Functioning Scale score (McPheeters, 1984; a measure of work, independent living and self care, immediate social network and extended social network), $F(3,31) = 3.19$, $p < .04$, with the adjusted $R^2$ indicating that these measures accounted for 16.2% of the variance. In contrast, out of a range of other measures of executive function not included in the BADS battery, only two; the Controlled Oral Word Association Test (Borkowski, Benton, & Spreen, 1967) and Porteus Mazes Test (Porteus, 1965) were able to discriminate between brain injured and neurologically healthy controls. However, neither of these was able to predict Role Functioning Scale score.

In sum, the multiple processes approach of the SAS is regarded here as being of most use to the study of executive functions. It is able to support the presence of a range of executive processes within one framework and several measures have been developed to test it which have ecological considerations built in. Accordingly, the measures selected for this study are those designed to test the SAS. Unfortunately, however not all measures of executive function have been developed with such ecological considerations in mind. Thus, prior to the measures used in this study being described more fully, a critical evaluation of measures that lack ecological considerations is offered.

## 1.6 Critical Evaluation of "Traditional" Measures of Executive Function

Whilst the benefits of experimentally testing theoretical models of executive function, are acknowledged, the leap from theory to practice is often complicated. Burgess and Robertson (2002) outline that there is often far less correspondence between the experimental paradigms used in research and the situations normally encountered by people in everyday life. Moreover, traditional tests of executive function that have been used to test such experimental paradigms often bear little resemblance to everyday tasks. The example of the MWCST, as described by Burgess et al. (2006), offers a useful illustration of this. Burgess et al. describe the MWCST as the most used and thoroughly investigated measure in the field of human frontal lobe function. The MWCST is typically considered a measure of “set shifting” where the examinee has to change the way they arrange a series of cards according to feedback from the examiner concerning whether or not they are pursuing the correct strategy. Burgess et al. suggest that, in theoretical terms, this test could plausibly be understood as a measure of working memory, supporting the ability to shift set. However, they suggest that as the MWCST is so unlike situations posed by everyday life, very few, if any, generalisations can be drawn about how MWCST performance may relate to everyday human activities or behaviours. Burgess and Robertson (2002) offer a similar argument for the aforementioned TOLT. As stated, the TOLT is traditionally considered a test of “planning” where the examinee is required to arrange a series of discs, onto a set of pegs, in a specified order, whilst obeying a series of rules. However, Burgess and Robertson suggest that planning in “real life”, which may take the form of arranging a weekend away, a holiday or a meal for friends is completely unlike the demands posed by the TOLT. Similar arguments can be levelled against most mainstream measures of executive function and neuropsychological assessments in general. This questions the ecological validity of most neuropsychological assessments. Accordingly, measures low in ecological validity challenge how accurate a clinician can be when developing a formulation of how a patient’s neuropsychological test scores may translate to their everyday behaviour.

This leads one to consider the utility of alternative assessments that can capture symptoms of the dysexecutive syndrome in contexts more relevant to everyday life. Indeed, Alderman and Baker (2009) describe such measures as having two advantages. Firstly, as they are designed to tap the cognitive demands that are required in everyday life, meaningful inferences can be drawn to assist in clinical assessment and formulation, to guide interventions relevant to “real life” impairments. Secondly, such assessments may also be useful tools in the actual remediation of executive difficulties themselves. However, prior to considering the availability of such measures it is first useful to discuss how the above discussion relates to people with ID.

## 1.7 Executive Function and Intellectual Disability: A Relevant Construct?

The vast majority of research investigating executive function has focused on people with brain injuries. Nevertheless, there also exist wide bodies of research examining cognitive impairment in people with dementia and schizophrenia, all of which have been the subject of considerable efforts to remediate such difficulties (Clare & Woods, 2001; Oddy & Worthington, 2009; Wykes & Reeder, 2005). Despite this, research examining executive function in the field of ID has been minimal. This is surprising because theoretical models of executive function may have much to offer in understanding the difficulties faced by people with ID in everyday life. Indeed, a diagnosis of an intellectual disability according to the DSM-IV-TR

(American Psychiatric Association, 2000) is considered to be (a) an IQ of approximately 70 or below and (b):

> "concurrent deficits or impairments in present adaptive functioning (i.e. the person's effectiveness in meeting the standards expected for his age or her age by his or her cultural group) in at least two of the following areas: communication, self-care, home living, social – interpersonal skills, use of community resources, self direction, functional academic skills, work, leisure and health and safety" (Carr & O'Reilly, 2007, pp. 7).

Looking at the above list, bearing in mind Oscar-Berman and Marinković's (2007) definition of executive function at the beginning of this chapter, it is plausible that deficits in executive function could serve as a useful framework through which impairments in, at least, home living, social – interpersonal skills and self direction could be formulated in people with ID, by making reference to models of executive functioning. Whilst the relationship between frontal lobe impairment and executive dysfunction may not be as straightforward in people with ID compared to people with overt frontal lobe damage, the argument advocated here, as outlined above, is that behavioural/functional observations are much more useful than anatomical ones for developing a clinical formulation of and intervention for dysexecutive symptoms.

Unfortunately, testing a possible executive function hypothesis in people with ID is difficult as the majority of standard measures of executive function tend to be too complex or rely too heavily on verbal skills for people with ID, resulting in floor effects (Masson, Dagnan, & Evans, 2010).

Nevertheless, recent research has attempted to adapt and evaluate measures of executive function for ID samples (e.g., Adams & Oliver, 2010; Ball, Holland, Treppner, Watson, & Huppert, 2008; Dymond, Bailey, Willner, & Parry, 2010; Lanfranchi et al., 2011; Masson et al., 2010; Willner, Bailey, Parry, & Dymond, 2010a). Of these, one measure which has appeared in all of the aforementioned studies is the TOLT. As outlined above, the TOLT was originally developed as a measure of the SAS. However, a major criticism of the TOLT is that, whilst it is largely regarded as a test of planning, it lacks ecological validity. This is true of nearly all of the other known measures of executive function that have been adapted for people with ID to date. Again, this limits the ability the clinician has to translate test scores onto the everyday behaviour of people with ID, thus limiting the implications that such assessments have for clinical formulation and intervention. Thus, there exists an opportunity to develop parallel measures of executive function for people with ID that are designed to have ecological validity. The potential for developing such measures is now discussed.

## 1.8 Developing Ecologically Valid Executive Function Assessments for People with Intellectual Disability

Willner et al. (2010a) suggest that one approach to test development in people with ID is to adapt versions of tests that have been developed for and validated on more able people. A brief search of the literature suggests that several measures which appear to have ecological validity do exist. Perhaps the most well known is the BADS described above. Nevertheless, whilst promising, the BADS remains an office based test and Willner et al. (2010a) found floor effects on three out of the six subtests from the children's version of the BADS (BADS-C; Emslie, Wilson, Burden, Nimmo-Smith, & Wilson, 2003) in people with ID, suggesting that this measure was at the lower limit of usability with ID groups. Accordingly, this test battery, whilst offering a more ecological approach to executive function assessment, may not be the most appropriate for people with ID.

Alternately, there exist several further ecologically based measures that could be considered as having the potential to be adapted for ID populations. For example, Baum et al. (2008) developed the Executive Function Performance Test. This comprises four individual tests including preparing a light meal, managing medications, using the telephone and paying bills. Baum et al. describe these tests as tapping the executive skills of initiation, execution, organisation, sequencing, judgement and safety and completion. Also, Chevignard et al. (2000, 2008) developed three tasks including shopping for groceries, cooking, answering a letter and finding a way to post the reply, all designed as measures of an individual's multitasking capabilities. Equally, Lamberts, Evans, and Spikman (2009) developed the Executive Secretarial Task, designed to test the organisation and prioritisation of multiple secretarial/administrative type tasks over a long time span, whilst dealing with delayed intentions, interruptions and deadlines. Overall, these measures clearly cover a range of everyday scenarios that are likely to challenge ones executive capabilities. However, the frequency with which they appear in the literature is limited to the few studies in which they are described meaning that the data on their psychometric properties, whilst promising, is limited. Nevertheless, one similar measure for which a much wider body of literature exists is the MET.

### 1.8.1 The Multiple Errands Test.

The MET was originally developed by Shallice and Burgess (1991) in order to capture the everyday task impairments of people with acquired brain injuries who were able to perform adequately on most mainstream neuropsychological tests, despite family members reporting significant impairments in their general activities of daily living. Accordingly, Shallice and Burgess developed a measure that could be undertaken within a real shopping centre in order to assess the examinee’s ability to function outside of structured office based settings. Here, the participants were given a list of tasks to undertake whilst following a series of rules. Whilst the tasks were relatively simple, the rules were designed to increase the test’s planning, multitasking and prospective memory demands (Alderman et al., 2003; Burgess & Alderman, 2004). Shallice and Burgess described successful performance on the MET as dependent on the ability of the SAS to (a) identify a goal, (b) create a plan, (c) create a “marker” to help the plan be realised effectively at a later time, (d) trigger the “marker” when necessary, and (e) monitor and evaluate the process to assist the creation of subgoals and/or modify of the plan if necessary. An example of the exercise sheet used in a version of the MET developed by Knight, Alderman, and Burgess (2002) is in Appendix A. Essentially, the participant is required to undertake six “doing tasks” (e.g., buy a can of Coca-Cola), four “information tasks” (e.g., find out the price of a Mars bar) and one “meeting” or prospective memory task (e.g., meet the examiner at a specified landmark after a certain period of time). Finally, the participant must tell the examiner when they have finished. The rules include tenets such as, “you should not go back into a building you have already been in”. The MET has four main outcome measures which are outlined in Table 1.2.

Table 1.2

*Examples of MET error categories on the MET*

| Error category | Definition | Example |
|---|---|---|
| Inefficiencies | Where a more efficient strategy could have been applied. | Buying multiple items or larger items than needed (e.g., buying more than four first class stamps or a bottle of coke rather than a can). |
| Rule breaks | Where a specific rule was broken. | Spent more money than allowed. |
| Interpretation failures | Where a task was misunderstood. | Writing down the opening and closing times of the library when only opening time was necessary. |
| Task failures | Where a task was not completed satisfactorily. | If the participant didn't meet the examiner within one minute of the meeting time or didn't state the time when meeting. |

*Note.* Adapted from "Beyond the shopping centre: using the Multiple Errands Test in the assessment and rehabilitation of multi-tasking disorders," By N. Alderman and D. Baker, In M. Oddy, and A. Worthington, 2009, *Rehabilitation of executive disorders: A guide to theory and practice,* p. 103. Copyright 2009 by Oxford University Press.

Shallice and Burgess (1991) originally found that three participants with frontal lobe damage who (as stated, despite performing in the average and above average ranges on most traditional neuropsychological tests, including those of executive function), performed much more poorly on the MET compared to controls. This work was supported by Goldstein et al. (1993) in a subsequent case study of a patient who underwent a left frontal lobectomy following a frontal lobe tumour who was found to generally perform in the average and above average range on most neuropsychological tests, but performed poorly on the MET.

These studies showed early promise in identifying the everyday impairments that can be overlooked by many neuropsychological tests. However, they were largely of small *n* or single case design and therefore relied primarily on descriptive statistics. The participants were also in the average to superior IQ ranges and it was therefore unclear how well people with lower IQs would cope with such a measure.

In response to this, Alderman et al. (2003) used a between groups design to investigate the utility of a simplified version of the MET adapted for shopping centres (MET-SV) that would avoid floor effects in people with lower IQs. Here, they assessed 50 people with brain injury and 46 neurologically healthy controls. Results showed that total MET-SV errors were significantly different between groups ($t = 4.03$, $p < .001$). MET-SV task failures correlated with BADS profile score ($r = -.46$, $p < .01$) and MWCST perseverative errors ($r = .39$, $p = < .01$), where high MET-SV scores and MWCST perseverative errors mean more executive dysfunction but a high BADS score means better executive function. Of 29 brain injured participants who also undertook three traditional executive tests used in this study, including the Cognitive Estimates Test (Shallice & Evans, 1978), a Verbal Fluency Test (Benton, 1968) and the MWCST, 17 passed them. However, only five of this 29 performed within acceptable limits on the MET-SV. This suggests that there are executive processes in the real world that may not be tapped by these traditional measures. In further analysing the MET-SV error types, Aldernman et al. found two factors emerged that characterised the brain injured participants' MET-ID performance, accounting for 62% of the variance. These included (a) rule breaking behaviour and (b) failure to achieve set tasks. These two categories showed significant differences in how independent raters scored them on the DEX questionnaire. Here, rule breakers were found to have more problems with executive memory symptoms (e.g., problems with confabulation and temporal sequencing), whereas task failures were found to show more negative affect (e.g., shallow affect and apathy).

Whilst the benefits of assessing people in shopping centres are clearly useful, this is often not pragmatically possible within the constraints of clinical settings. Knight et al. (2002) therefore used a between groups design to investigate the utility of an adapted version of the MET for use in a hospital grounds (MET-HV) to ease the difficulties that might come about from assessing participants in more public settings. Here, they assessed 20 participants with brain injury and 20 neurologically healthy controls. They found a significant difference between MET-HV total errors across groups ($t = 5.25$, $p < .001$). MET-HV total errors also correlated highly with BADS total profile score ($r = -.57$, $p = .009$) and MWCST perseverative errors ($r = .67$, $p = .001$), where high MET-HV scores and MWCST perseverative errors mean more executive dysfunction but a high BADS score means better executive function. After partialling out age, memory and familiarity with hospital grounds, MET-HV task failures correlated with four of the five DEX outcome measures (between $r = .51$ and .79, all $p < .038$), where both high MET-HV and DEX scores mean worse executive function. This work was supported by Dawson et al. (2009) who extended the MET-HV to assess 14 post stroke participants, 14 participants with a brain injury and 25 neurologically healthy controls on the measure. The effect size between MET-HV performance of the stroke participants and their controls was large ($d$ = .98) whereas the effect size between the brain injury group and their controls was moderate ($d$ = .54). Inter-rater reliability across outcome measures ranged from $r$ = .71 to .88 (all $p < .001$). In further work that was designed to overcome difficulties associated with assessing patients who might not be able to leave the confines of a hospital ward (for legal or pragmatic reasons), Pennington (2006) used a between groups design investigating the utility of a version of the MET adapted for a hospital ward (MET-WV) in 21 participants with brain injury and 24 neurologically healthy controls. It was found that total MET-WV errors were significantly different between groups ($t$ = 5.23, $p < .001$) and significantly related to BADS profile score ($r$ = -.55, $p < .05$), where high MET-WV scores mean worse executive function but high BADS score means better executive function.

Other work has taken this a step further and adapted the MET for use in virtual environments. Here, McGeorge et al. (2001) developed a virtual reality version of the MET for a university department. They assessed five participants with brain injury and five neurologically healthy controls, finding a significant difference between the number of errands completed by the two groups ($F$ = 8.86, $p < .05$). The correlation between performance on the virtual MET and its real life university department analogue across all participants was high ($r$ = .79, $p < .01$). Subsequently, Rand, Basha-Abu, Rukan, Weiss, & Katz (2009) developed a virtual reality version of the MET-

SV and compared nine post stroke participants, 20 neurologically healthy "young" participants and 20 neurologically healthy "older" participants. There was a strong correlation between total errors on an adapted version of the MET-SV and the virtual MET-SV ($r = .77$, $p < .01$). A significant correlation also emerged between virtual MET-SV total errors and performance on the Zoo Map subtest from the BADS ($r = -.93$, $p < .001$) and the Instrumental Activities of Daily Living Questionnaire ($r = -.76$, $p < .04$) in the post stroke group, where high scores on the virtual MET-SV mean worse executive function but high scores on the Zoo Map subtest and Instrumental Activities of Daily Living Questionnaire mean better functioning. Virtual MET-SV errors were also higher in the post stroke group, followed by the neurologically healthy "older" participants, whose errors were greater than the neurologically healthy "younger" participants ($\chi^2 = 30.07$, $p < .001$). A difficulty with these studies however is the small samples used. In addition, it is unclear the extent to which virtual environments can be considered representative of real world environments as they are arguably far removed from a direct analogue of everyday living. Nevertheless, such virtual environments are pragmatic and perhaps of most importance have given rise to work examining rehabilitative interventions. Accordingly, Rand, Weiss, & Katz, (2009) report a case series to investigate the utility of the virtual MET-SV as a tool remediating multitasking impairments post stroke. Here, they gave four post stroke participants ten, 60 minute training sessions on the virtual MET-SV over a 3 week period. Descriptive statistics showed that all participants demonstrated an improvement from pre to post intervention performance on the real life version of the MET-SV after training on the virtual MET-SV. However, none was able to attain a fully independent level of shopping post intervention.

A difficulty with the aforementioned studies however is that despite many of the participants being taken from a heterogeneous group of “brain injured” or post stroke participants, neuroanatomical issues have been largely neglected in MET research. Whilst the limitations of this approach have been considered above (see section 1.3), based on research that damage to the ventromedial prefrontal cortex (VMPC) is a region of the frontal lobes believed to be heavily implicated in real world decision making (Bechara & Damiaso, 2005; Eslinger & Damasio, 1985), Tranel, Hathaway-Nepple, and Anderson (2007) investigated whether damage to this area of the brain would predict impaired performance on the MET-SV relative to brain damage outside of this region. Accordingly, they assessed nine participants with bilateral damage to the VMPC, 8 with frontal damage outside the VMPC, 17 with non frontal brain damage and 20 neurologically healthy controls. A one way analysis of variance revealed a main effect of group for MET-SV total errors ($F$ [3, 50] = 6.78, $p$ = .001), where the VMPC group had the largest number of errors, those with frontal damage outside the VMPC region attaining the second largest number of errors, those with non frontal brain damage scoring the third and the healthy controls making the fewest errors.

Further work attesting to the involvement of the neuroanatomical regions involved in the MET is offered by Burgess, Alderman, Volle, Benoit, and Gilbert (2009) who presented data from the neuropsychological assessments of two participants who took part in the Alderman et al. (2003).

The authors suggest that the data support a double dissociation where, contrary to previous findings of above average neuropsychological test scores and impaired MET performance, the two participants showed normal MET-SV performance in the presence of impaired performance on other neuropsychological tests. From this, Burgess et al. suggest that there are brain areas that support MET performance which are independent of those used in tests of IQ, memory, or other executive function abilities.

Whilst the data is limited to brain injured and post stroke populations, and the MET remains under relative development, its strengths lie in its inherent ecological validity as a tool for assessing functional difficulties, which may be used to inform targeted remediating interventions. It is undertaken in an environment outside of the typical office setting and can provide a quantitative and seemingly ecologically valid measure of executive function in the real world, as well as give the examiner a subjective understanding of the participant’s ability to deploy executive skills in everyday life. This may therefore be a useful measure to adapt for the ecological assessment of executive function in people with ID.

## 1.9 Review of the Argument, Aims of the Study and Research Hypotheses

This introduction section has outlined the argument for the need to develop ecologically valid assessments for people with ID. Initially, it was suggested that a focus on the behavioural symptoms of executive dysfunction (e.g., the dysexecutive syndrome) may serve as more effective means of informing clinical formulation and intervention, above and beyond the practice of anatomical localisation. It was then suggested that theoretical models of executive function (particularly the SAS) may help the clinician understand some of the everyday difficulties faced by people with ID, and provide implications for remediating such difficulties. However, the traditional measures with which executive function has been assessed were criticised for lacking ecological validity, including the executive function measures that have been adapted for people with ID. The MET was therefore then suggested as a potential measure that could be adapted for people with ID in order to provide an assessment of how executive function deficits may play out in the everyday life of someone with ID.

The aims of the study were therefore to assess whether a modified version of the MET for people with ID (MET-ID) could be developed. The ecological validity of the MET-ID was then assessed by correlating MET-ID performance with three measures of executive function (or dysexecutive symptoms). Traditionally, formal means of assessing the ecological validity of neuropsychological tests has taken the form of exploring how well self and informant based questionnaires, clinician rating scales and observation of everyday simulated tasks correlate with neuropsychological test performance (Chaytor, Schmitter-Edgecombe, & Burr, 2006). Accordingly, the Independent rater version of the DEX (DEX-IR; Burgess et al., 1996) was used, alongside a version of the TOLT adapted for people with ID (Masson et al., 2010) and the Six Parts Test (SPT; Emslie et al., 2003). The SPT is a subtest from the BADS-C, and a simplified version of the SET. As stated above, Shallice and Burgess (1991) first developed the SET along with the

MET, as an office based means of capturing the everyday task impairments that appeared to be overlooked by traditional neuropsychological assessments (Burgess, 2000). The SPT thus offers a plausibly ecologically valid, but office based measure to compare with MET-ID performance, alongside the purportedly less ecologically valid TOLT. The research hypotheses are therefore:

1) The DEX-IR will correlate more strongly with the MET-ID scales and the SPT compared to the TOLT as the MED-ID and SPT are designed to be more predictive of the “real life” symptoms of SAS impairment.

2) There will be a stronger correlation between the MET-ID scales and the SPT compared to the correlation between the MET-ID and the TOLT as the MET-ID and SPT are designed to tap a subset of the same cognitive components required in “real life” tasks.

# Chapter Two: Method

## 2.1 Overview

This chapter outlines the methodology used in the study. Firstly, the design is outlined before details about the participants in terms of inclusion/exclusion criteria, recruitment, power analysis and ethical considerations are presented. The measures used in the study are then described along with details of their administration scoring, validity and reliability. The section closes by outlining the procedure and plan for analysis.

## 2.2 Design

The study employed a cross sectional, correlational design using a single sample of participants with ID. This was appropriate so that measures of the SAS could be compared to assess their degree of correlation. Accordingly, there was only one group of participants measured on one occasion.

Previous investigations using the MET have employed between group designs where the MET performance of groups with potential executive impairment (e.g., those with acquired brain injuries) has been compared with the MET performance of neurologically healthy controls. This was considered inappropriate for the present study as it was anticipated that the MET-ID would require substantial adaptation to make it accessible to an ID sample. Such adaptation may therefore compromise the MET-ID’s validity in non ID populations making group comparisons difficult. Therefore, the aims of this study were simply to establish a basis for the utility the MET-ID in an ID sample and compare it with other measures of the SAS.

## 2.3 Participants

### 2.3.1 Inclusion/exclusion criteria and demographic data.

Inclusion criteria were adults, aged 18-65, attending day centres for people with ID. Participants were required to have an IQ within the 50-70 range as measured by the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) to indicate the presence of an ID of “mild” severity (American Psychiatric Association, 2000). Exclusion criteria were those who had an acquired brain injury, dementia or schizophrenia as these can impair the SAS independent of ID (Marczewski et al., 2001; Perry & Hodges, 1999; Wilson et al., 1996).

Overall, 48 participants took part. This included eight participants used to pilot a provisional version of the MET-ID and 40 participants used to trial the final adapted version (details of the piloting procedure are given below). The eventual sample comprised 62.5% females, with a mean (with standard deviations in parentheses) age of 45 (9.27) years, range 28-64. The sample’s mean IQ score was 58 (4.54), range 52-70.

### 2.3.2 Recruitment.

Participants were recruited from day centres for people with ID in the Norfolk region. As it was recognised that people with ID may be vulnerable to acquiescence (which may make them liable to consent to participate without understanding the full implications) the manager of the day centre was asked to identify potential participants for the study, as opposed to the chief investigator (CI) approaching them directly. Specifically, the day centre manager was asked to identify participants who were felt able to understand (a) the purpose and nature of the research, (b) what the research involves, (c) its benefits (or lack of benefits), (d) the risks, burdens and alternatives to taking part, (e) be able to retain the information long enough to make an effective decision, (f) be able to make a free choice and (g) be capable of making the particular decision at the time it needs to be made. On top of this the manager was asked to exclude anyone who might have dementia, schizophrenia and/or acquired brain injury.

Once identified, a member of staff well known to the potential participant was asked to assess the participant's interest in taking part by introducing them to an information sheet (Appendix B). This was again to reduce potential pressure on the participant to consent. Only if verbal consent was initially given did the CI then approach the participant, reintroduce the information sheet, seek written informed consent (Appendix C) and collect demographic data (Appendix D).

The study also required an "informant" (e.g., member of day centre staff) to complete the DEX-IR for each participant. Here, staff members were treated as participants in their own right and were given information about the study by the day centre manager who subsequently asked if they would volunteer to give ratings. Those who responded favourably were given a separate information sheet (Appendix E) and subsequently met with the CI to provide informed consent (Appendix F).

#### 2.3.3 Sample size.

At time of writing, there was no known accepted method for calculating power in situations where correlated correlations will be compared. In view of this, a sample size was estimated using a method outlined by Bush (2011), based on what was judged to be an acceptable lower bound 95% confidence interval surrounding a desired correlation between the MET-ID and the DEX-IR. This method was adopted because it allows an estimate of the number of participants required to achieve a specified correlation between two variables, based on a parameter for how far any correlation falling below the lower bound confidence interval would need to be in order for it to be considered unlikely to occur in the general population. In previous work examining correlations between the MET and observer ratings of executive dysfunction in everyday life, Knight et al. (2002) found that MET task failures correlated with four out of the five factor scales on the DEX-IR; between $r = .51$ and $r = .79$ (all $p < .05$). Therefore, a desired r of .80 was adopted for this study. In order to calculate the 95% lower bound confidence interval around this, a sample size of 40 was chosen, based on previous related work where this sample size has been used (e.g., Dymond et al., 2010; Masson et al., 2010; Willner et al., 2010a). Fisher's r to z transformation was then applied to convert r into a figure that is approximately normally distributed (e.g., Fisher's z), which has a standard error of $1/\sqrt{(n-3)}$. Accordingly, an r value of .80 gives a Fisher's z value of 1.0986. The standard error of Fisher's z, using a sample size of 40 is .164. Using a z score of 1.64 for a one sided prediction, the lower limit confidence interval would therefore be 1.0986 - (1.64 x 0.164) which gives a Fisher's z score of 0.8362. When converted to r this gives a lower limit confidence interval of .68. This was deemed an acceptable lower bound 95% confidence interval and therefore the study aimed to recruit 40 participants.

#### 2.3.4 Ethical considerations.

Ethical considerations were reviewed in accordance with the guidelines for minimum standards of ethical approval in psychological research (British Psychological Society, 2004). To avoid potential distress caused by task failure, four of the measures incorporated discontinue criteria so that continued failure resulted in test cessation. For the two measures with no discontinue criteria, the participant could stop at any time as successful performance was self determined. To avoid further distress through potentially overloading the participants, the assessments took place over two, 1 hour sessions.

Participants were also asked for their consent to a member of staff completing the observer rating of executive dysfunction and made aware of their right to withdraw via the consent form. They were also asked to indicate after each session, “I liked it a lot”, “I liked it” or “I didn’t like it”. It was planned that any participant who indicated ‘I didn’t like it’ would be reminded of their right to withdraw and their key worker would be informed. Of the 50 first sessions, 94% were reported as “I liked it a lot”, 6% were reported as “I liked it a bit” and 0% were reported as “I didn’t like it”. Of the 50 second sessions, 93.9% were reported as “I liked it a lot”, 6.1% were reported as “I liked it a bit” and 0% were reported as “I didn’t like it”.

All data were anonymised and immediately transferred to a statistical software spreadsheet and saved on a password protected laptop. The laptop and the raw data were kept in a locked draw during the data collection period under requirements of the Data Protection Act (1998). Ethical approval was obtained from Cambridge 4 Research Ethics Committee, reference number 10/H0305/75 (Appendix G) and appropriate research governance approval was granted by NHS Norfolk ref 2010LD02 (Appendix H).

## 2.4 Measures

### 2.4.1 Supervisory attentional system.

Three measures of the supervisory attentional system were employed. These included the MET-ID, a version of the TOLT adapted for people with ID and the SPT.

#### ***2.4.1.1 The Multiple Errands Test.***

##### *2.4.1.1.1 Description.*

The MET has been described in section 1.8.1. Accordingly, it was considered necessary to adapt the measure for an ID population. Details of the adaptation, piloting and scoring of the MET-ID are outlined below.

##### *2.4.1.1.2 Adaptation, pilot testing and scoring of the MET-ID.*

An initial version of the MET-ID was developed for pilot testing (Appendix I). This was modelled as closely as possible to the Knight et al. (2002) and Pennington (2006) versions but modified in a number of ways. Firstly, the text was enlarged, simplified and spread across an A3 sheet with pictures added to supplement the written material. Secondly, to assist those with reading difficulties, auditory equipment was used to vocalize the written instructions. Here, a “recorderPEN”[1] was adopted where the participant could move a “pen” over small labels placed next to text on the MET-ID exercise sheet, prompting the pen to read out the tasks/rules. Thirdly, the tasks were divided into green “doing tasks”, blue “information tasks” and a red “prospective memory task” to help them be easily differentiated. Fourthly, the green tasks (doing tasks) were reduced from six, as per Knight et al. to three, and the blue tasks (information tasks) were reduced from four to three. The rules were also reduced from nine to seven. Fifthly, participants were required to tell the examiner the information in the “information tasks” as

[1] Purchased from http://www.mantralingua.com/product.php?productid=16135&js=y

opposed to writing it down. This initial adaptation was piloted using eight participants. Data from the eight participants can be seen in Table 2.1.

Table 2.1

*Pilot data from eight participants undertaking the initial version of the MET-ID*

| Participant number | Site | VIQ | PIQ | FSIQ | BPVS | WR | MET-ID |
|---|---|---|---|---|---|---|---|
| 1 | A | 72 | 64 | 66 | 115 | 59 | Initiated tasks |
| 2 | A | 68 | 55 | 59 | 118 | 54 | Initiated tasks |
| 3 | A | 63 | 58 | 57 | 111 | 64 | Initiated tasks |
| 4 | B | 57 | 55 | 53 | 65 | 3 | Initiated tasks |
| 5 | B | 59 | 55 | 54 | 81 | 55 | FT |
| 6 | B | 67 | 55 | 59 | 90 | 48 | Initiated tasks |
| 7 | B | 63 | 57 | 57 | 108 | 90 | Initiated tasks |
| 8 | B | 65 | 55 | 57 | 66 | 33 | FT |

*Note.* BPVS = British Picture Vocabulary Scale – Third Edition; FSIQ = Wechsler Abbreviated Scale of Intelligence Full Scale IQ; FT = Failed to attempt any tasks; MET-ID = Multiple Errands Test – Intellectual Disability Version; PIQ = Wechsler Abbreviated Scale of Intelligence Performance IQ; VIQ = Wechsler Abbreviated Scale of Intelligence Verbal IQ; WR = Wechsler Individual Attainment Test Word Reading subtest raw score.

As can be seen in Table 2.1 participants five and eight failed to initiate any of the tasks. Analysis of these participants' performance showed that they were able to use the pen to read out the tasks however they didn't actually attempt any of them. Rather, they remained stationary and when they had finished scrolling through the tasks looked at the examiner. Otherwise, participants one, two, three, four, six and seven attempted the

MET-ID. Analysis of these participants performance showed that all of these participants attempted the green doing tasks.

However, the blue information finding tasks were only attempted by participants two, four, five and seven. Moreover, participants four and seven attempted the blue information tasks without actually going on any errands. For example, they estimated the number of clocks at the day centre and told the examiner one of the day's activities and the date without actively seeking this information. Thus, these tasks were plausibly failing to tap the desired aspects of these participants' executive skills. Another issue was that successfully completing some of the information tasks may have actually been dependent upon reading ability in its own right (e.g., reading the date and afternoon's activity from an activities board) rather than executive function.

Equally, none of the participants completed the red prospective memory task of "Tell Tom when five minutes has passed on the watch". This was plausibly down to many participants finishing the MET-ID in under five minutes, but also the fact that some of the participants had difficulties telling the time. In addition, seven of the eight pilot participants broke the rule, "don't speak to Tom unless it's for one of the tasks" and all the participants broke the rule "Do all the tasks, but in any order."

These findings were discussed with the author's research supervisors (both Clinical Psychologists) and the MET-ID was modified for a second time (Appendix J). Here, the blue information finding tasks were removed and replaced with three further green doing tasks. The red meeting or prospective memory task was incorporated into one of these and changed to “clap your hands together after three minutes.” Successful completion of this task was considered to be clapping hands together any time in between two and four minutes. Also, in view of the information tasks being removed, the rule “don’t speak to Tom unless it’s for one of the tasks” was simplified to “don’t speak to Tom unless it’s to tell him when you have finished.” Equally, to simplify the measure further, the rules “don’t carry more than one thing” and “don’t rush” were dropped as it was considered to be unlikely that the participants would carry more than one thing and difficult to objectively quantify whether a participant is rushing. The tasks were also coloured as green and the rules as red to help them be further differentiated.

The second version was administered to five new participants that had not attempted the original version by initially explaining the task as outlined in Appendix K. As part of this process, participants were asked to revise the rules (with the assistance of the auditory equipment) over one minute and then recall them without looking at the exercise sheet. The number of rules recalled correctly at first attempt was recorded. Any unrecalled rules were subsequently prompted by the examiner and again recorded (see Appendix L). The participant was then reminded of any rules left unrecalled. The purpose of this was not a test of memory, but to ensure that the participant knew the rules and familiarised them with the auditory equipment.

Likewise, in order to try and overcome the fact that some participants were unable to initiate any of the tasks, two practice tasks were introduced prior to the full administration of the MET-ID (Appendix M). The participants were guided through these in order to help them understand the fact that they had to actually initiate the tasks, rather than just scroll through them with the pen. When the examiner felt that the participant had a clear grasp of the instructions they were asked to begin the exercise.

Once started, the examiner followed the participant around the day centre at a distance of approximately two metres until the participant informed the examiner that they had finished. Participants carried the MET-ID instructions and the auditory recorder pen with them whilst they were carrying out the task. Finally, a standard prompt was introduced when a participant asked a question or tried to speak to the examiner, in order to stop any covert support that might be inadvertently provided. This was: “remember, I’m not allowed to give you any help but you must do all the tasks and not break any of these rules”.

This second version of the MET-ID was analyzed using the Flesch-Kincaid readability test on Microsoft Word 2007. This returned a Flesch-Kincaid grade level of 1.9, which means that a typically developing child in their second grade of the United States education system, who would typically be seven or eight years of age, should be able to read the MET-ID instructions. Five participants’ performance on the second version is outlined in Table 2.2.

Table 2.2

*Pilot data from five participants undertaking the second version of the MET-ID*

| Participant number | Site | VIQ | PIQ | FSIQ | BPVS | WR | MET-ID |
|---|---|---|---|---|---|---|---|
| 9 | B | 62 | 55 | 55 | 63 | 8 | FT |
| 10 | B | 72 | 57 | 62 | 128 | 108 | Initiated tasks |
| 11 | B | 57 | 55 | 54 | 59 | 24 | Initiated tasks |
| 12 | B | 55 | 55 | 52 | 92 | 51 | Initiated tasks |
| 13 | B | 71 | 59 | 63 | 131 | 95 | Initiated tasks |

*Note.* BPVS = British Picture Vocabulary Scale – Third Edition; FSIQ = Wechsler Abbreviated Scale of Intelligence Full Scale IQ; FT = Failed to attempt any tasks; MET-ID = Multiple Errands Test – Intellectual Disability Version; PIQ = Wechsler Abbreviated Scale of Intelligence Performance IQ; VIQ = Wechsler Abbreviated Scale of Intelligence Verbal IQ; WR = Wechsler Individual Attainment Test Word Reading subtest raw score.

As can be seen in Table 2.2 participant nine was the only one unable to initiate any of the MET-ID tasks. Her performance showed the same pattern of scrolling through the tasks but not actually initiating any of them, despite being guided through two practice tasks with the examiner only moments earlier. This didn’t appear to be related to any comprehension difficulties as her WASI verbal IQ and BPVS scores were actually higher than two of the other participants who were able to engage in the MET-ID tasks. However, her WIAT score was the lowest of all five suggesting a low level of reading ability that could plausibly explain her inability to initiate tasks. Nevertheless, this final version of the MET-ID was adopted so the task initiation phenomenon could be investigated further. Specifically, it was unclear whether the inability to initiate any tasks was down to a failure in task comprehension or an executive function deficit. This is considered in section 3.5 below. Data from these five participants were used in the final analysis.

Recording of MET-ID performance was conducted as per Knight et al. (2002) where the examiner wrote down all aspects of the participants' performance as they engaged in the task. An example of this is given in Appendix N. This written summary of performance was then scored. In developing a scoring system, interpretation failures and inefficiencies were abandoned for three reasons. Firstly, it was felt that, if a participant fails to actually initiate any tasks, they are less likely to break rules, make interpretation failures and commit inefficiencies. For example, pilot testing showed that many participants were giving the examiner a magazine rather than a book. This could have been coded as a task failure (where a task was not completed satisfactorily) and an interpretation failure (where a task was misunderstood). However, for those who didn't attempt this task at all, this would have just been coded as a task failure. Therefore, the participants giving the examiner a magazine rather than a book would have received a higher error profile score (meaning worse executive function) than the participants who didn't actually initiate the task at all. Secondly, it also proved hard to actually define and code instances of interpretation failures and inefficiencies before the MET-ID was actually trialled. Indeed, it appears that these measures could have only been defined and coded on a retrospective basis. As this may compromise the validity of any such recordings, no attempt was made to quantify them. Thirdly, interpretation failures and inefficiencies have been found to occur vastly less frequently than task failures and rule breaks (Alderman et al., 2003).

In view of this, two alternative scales were introduced. These included task attempts (where a task is attempted but not completed satisfactorily) and task completions (where a task is attempted and completed satisfactorily). Here, the task attempts category was included in order to give credit for task initiation, even if the task was performed incorrectly. A similar scale crediting task attempts was employed by Dawson et al. (2009) on their version of the MET-HV. Alternately, task completions were designed to credit both successful attempts and completions. Task completions also stood as a "mirror image" to task failures from the original MET, so could replace this scale. Both these scales were scored out of six, with one point being awarded for each of the six tasks. In keeping with the original MET, rule breaks were also recorded and scored out of six where a point was awarded if a rule was broken. The system used to score each of the three scales is outlined in Appendix O.

##### *2.4.1.1.3 Validity.*

As the MET has not been trialed before in an ID population, the data reported here are taken from its use with different samples. Whilst some of these data have been described in section 1.8.1 above, several statistics are repeated here for illustration purposes.

In comparing MET performance across clinical groups, Knight et al. (2002) found that total error score on a hospital based version of the MET discriminated between participants with acquired brain injuries and "healthy" controls ($t$ = 5.25, $p$ < .001). This was also found by Alderman et al. (2003, $t$

= 7.18, $p < .001$) and Pennington (2006, $t = 5.23$, $p < .001$) on shopping centre and hospital ward versions of the MET respectively.

Moreover, in terms of ecological validity, as the MET is undertaken in a real life setting as opposed to the laboratory, it has potentially useful ecological considerations built in (see Alderman et al., 2003; Burgess et al., 2006; Dawson et al., 2009). Accordingly, the MET has been frequently compared with the DEX-IR to investigate how it relates to everyday symptoms of executive dysfunction. Here, in samples with acquired brain injuries, DEX-IR total score has correlated with MET errors, ($r = .69$, $p < .05$, Dawson et al., 2009) and MET task failures, ($r = .79$, $p < .001$, Knight et al., 2002), respectively, where high scores on the DEX-IR and MET indicate more executive dysfunction. Furthermore, Dawson et al. (2009) found that MET performance was related to a standardized assessment of activities of daily living; the Assessment of Motor and Process Skills (AMPS; Fisher & Bray Jones, 2010), where AMPS process score correlated with MET rule breaks, ($r = -.69$, $p < .05$) and MET errors ($r = -.38$, *ns*), with higher AMPS score relating to better performance on activities of daily living.

In terms of convergent validity, the MET has been largely correlated with the BADS, as this also purports to be an ecologically sensitive measure of executive function. Accordingly, MET total errors have correlated with BADS total profile score, ($r = -.57$, $p < .009$, Knight et al., 2002; $r = -.46$, $p < .01$, Alderman et al., 2003; $r = -.55$, $p < .05$, Pennington, 2006) on hospital grounds, shopping centre and ward based versions of the MET respectively, where lower BADS scores relate to more executive dysfunction. In examining the individual BADS subtests, MET total errors have correlated with the

BADS Zoo Map test profile score, ($r_s$ = -.93, $p$ < .001, Rand, Basha-Abu Rukan, Weiss, & Katz, 2009; $r$ = -.46, $p$ < .01, Alderman et al., 2003; $r$ = -.41, *ns*, Knight et al., 2002), as well as the BADS SET profile score ($r$ = -.41, $p$ < .01, Alderman et al., 2003) in those with acquired brain injuries. The MET also appears to have a relationship with card sorting tests, which, whilst making no claims to ecological validity, have themselves been shown to be sensitive to frontal lobe lesions (Demakis, 2003). Here, MET total errors have correlated with MWCST perseverative errors ($r$ = .67, $p$ < .001, Knight et al., 2002; $r$ = .39, $p$ < .01, Alderman et al., 2003), where higher MWCST perseverative errors mean more executive dysfunction.

*2.4.1.1.5 Reliability.*

Knight et al. (2002) obtained inter-rater reliability scores ranging from .81 to 1.00 and Dawson et al. (2009) found similar inter-rater reliability scores of between .71 and .88 for the different outcome measures of the MET. Knight et al. also found Cronbach’s alpha internal consistency reliability of 0.77 for each outcome measure’s ability to predict MET total score. Rand et al. (2009) also demonstrated alternate form reliability between virtual reality and “real life” shopping centre versions of the MET, showing a significant correlation in total MET errors on both versions across a total sample of 49 post stroke and healthy participants ($r$ = .77, $p$ < .01).

#### *2.4.1.2 The Six Part test.*

##### *2.4.1.2.1 Description.*

As outlined above, the SPT is a subtest from the BADS-C and a simplified version of the SET. Whilst the SET is undertaken in the "laboratory", it is designed to "tap a subset of the same cognitive components" (Burgess, 2000, p. 281) that are required in the MET and everyday situations. Indeed, Burgess (2000) describes how many everyday situations such as shopping or cooking have the following features; many discrete tasks, merging of these tasks is necessary for efficient performance, only one task can be performed at a time (due to cognitive or physical constraints), unforeseen interruptions may occur, delayed intentions may need to be activated, the component tasks have differing characteristics, acceptable performance is self determined and no immediate feedback is given (e.g., failures are not immediately highlighted). Burgess explains how the SET contains all of the characteristics outlined above (with the exception of potential unforeseen interruptions) and that these characteristics tend to be overlooked by the majority of other neuropsychological measures. Specifically, the SPT has been described as a measure of "Planning, task scheduling and performance monitoring" (Baron, 2006, p. 540) and is described below.

The SPT contains three tasks, each with two parts (six parts in total). The first task: "How many?" includes two booklets (parts one and two) with each page featuring a different number of items. The participant is required to state the number of items in each of the pictures. The second task: "What is it?" also includes two booklets (parts one and two) with a different single object on each page. The participant is required to name the object in each picture. The third task: “Sort me” contains two boxes (parts one and two) each containing small objects. The participant has to put a specified type of small object into the lid of each respective box as indicted on the upturned side of the box’s lid. The test requires that the participant must attempt at least something from each of the six parts within a five minute time limit. However, a rule prevents the participant from completing two parts from the same task consecutively. For example, the participant cannot go directly from “How many?” part one to “How many?” part two without first attempting a part of the “What is it?” or “Sort me” tasks. The participant has a visible countdown timer to help them keep track of the time and a written rule sheet outlined in front of them.

##### *2.4.1.2.2 Administration and scoring.*

The SPT was administered as per the BADS-C manual, but modified to allow the participant to read out the answers to the “How many?” and “What is it?” tasks, as opposed to writing them down. The rule sheet was also enlarged and supplemented with pictures to assist those with reading difficulties.

The maximum score available on the STP is 16. Points are scored by the number of tasks attempted and the strategies used by the participant to complete the task. Points can be lost if a participant breaks a rule on one of the tasks (e.g., does “How many” part one followed by “How many” part two, without attempting one of the “What is it” or “Sort me” tasks in between). Lower score means worse executive function.

*2.4.1.2.3 Validity and reliability.*

Due to the relative dearth of available tests of EF in people with ID, there are no known data attesting to the validity or reliability of the SPT in an ID sample. However, the SPT was investigated by Willner et al. (2010a) as part of an investigation into the utility of measures of executive function in 40 people with ID. They found that no participant scored at floor or ceiling level on the SPT ($M$ = 3.5, $SD$ = 1.6). The SPT also showed a medium, significant correlation with the TOLT, ($r$ = .344, $p$ < .05; P Willner, personal communication, July 3, 2010), where high scores on the TOLT and SPT indicated better performance.

The validity of the SPT when tested by Emsile et al. (2003) indicated a negative correlation between the SPT and a version of the DEX-IR modified for children in their combined neurologically healthy and clinical sample of 314 children ($r$ = -.208, $p$ < .000), where higher DEX-IR score means worse executive function. The SPT also showed a significant negative correlation ($r$ = -.275, $p$ < .000) with the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1999) total difficulties score, where higher SDQ score means more difficulties

Emslie et al. (2003) also demonstrated inter-rater reliability for the SPT in a sample of 25 neurologically healthy participants aged 8 to 14 where correlations between raters for the number of tasks attempted, number of rule breaks and total score ranged from .92 to 1.00 (all $p < .001$). Emslie et al. also assessed the test retest reliability of the SPT in this sample. As a critical aspect of tests of executive function is novelty, since it is new situations that people with executive problems are likely to find difficult, it was not expected that there would be strong test-retest reliability. Indeed, the samples mean score was significantly lower when re-tested three to four weeks after the original administration ($p < .006$). This is in line with findings of other tests of frontal lobe functioning, including the adult version of the BADS (Emslie et al., 2003).

Whilst the SPT's validity and reliability in an ID population is unestablished, the SPT is one of the few tests available that aims to tap the skills required in the MET. To this end it stands as a useful laboratory based measure with which to correlate MET-ID performance.

#### *2.4.1.3 The Tower of London Test.*

##### *2.4.1.3.1 Description.*

The TOLT was originally developed by Shallice (1982) as a measure of the planning processes governed by the SAS (Andres & Van der Linden, 2000, 2001; Ball et al., 2008; Chan et al., 2008; Marczewski, Van der Linden, & Larøi, 2001; Masson, et al., 2010; Rainville et al., 2002). Subsequently, the TOLT (or variants of it) have been incorporated into executive function assessment batteries such as the Delis-Kaplan Executive Function System (Delis, Kaplan, & Krammer, 2000) and the Cambridge Executive Function

Assessment (CEFA; Ball et al., 2008). As a description of the TOLT has been given in the introduction section, it will not be repeated here.

#### *2.4.1.3.2 Administration and scoring.*

The TOLT administration and scoring procedures as developed by Masson et al. (2010) in a study investigating the utility of the TOLT in an ID sample were used. The participants were initially asked to count the six discs to ensure that they could count to six as this was the maximum number of moves they would be required to use. A motor co-ordination task was then initially modelled by the examiner before asking the participants to place all six discs onto the pegs in any order. The rules were then explained and a two move problem was modelled by the test administrator followed by a practice two move problem for each participant.

The test started with a first problem that required one move, a second problem that required two moves, and so forth, and ended with a sixth problem that required six moves. A time limit was placed on the one and two move problems of 30 seconds, the three and four move problems of 60 seconds and the five and six move problems of 120 seconds. However, this was for pragmatic purposes, as emphasis was on accuracy rather than speed. Each time a rule was broken the trial was stopped, the rule break was pointed out and the participants were reminded of the rules. The participant was allowed three trials per problem.

A total points score (out of 18) was awarded where participants would get three points if they passed a problem on the first attempt, two if passed at second attempt, one if passed at third and zero if not passed at third attempt. The number of problems solved correctly at first attempt (out of six) was also recorded, as well as the highest problem level achieved (out of six). Lower scores mean worse executive function. A discontinue rule was applied if any participant failed all three trials of two consecutive items.

*2.4.1.3.3 Validity.*

As outlined above, performance on the TOLT has been shown to differentiate between “healthy” controls and participants with dementia (Rainville et al., 2002), schizophrenia (Marczewski et al., 2001) and acquired brain injuries (Shallice, 1982). Successful performance on the TOLT has also been shown to be impaired in “older” (relative to “younger”) groups (Andres & Van der Linden, 2000). The TOLT has also been shown to discriminate between those mild and moderate ID who have Down’s Syndrome (Ball et al., 2008).

Masson et al. (2010) found a clear hierarchical structure where all 43 participants with ID undertaking the TOLT were able to solve the first problem level but only 9 (20.9%) were able to solve the sixth and final level. Unfortunately, Masson et al. did not include any other “laboratory” based measures of executive function so convergent validity could not be assessed. In terms of ecological validity, Masson et al. (2002) compared TOLT performance with the DEX-IR and the Adaptive Behaviour Scale – Residential and Community: Second Edition (ABS-RC 2; Nihira et al., 1993). Higher ABS-RC 2 scores mean better adaptive functioning. They found that the DEX-IR correlated with TOLT total score (*r* = -.507, *p* = .001), TOLT correct first attempts, (*r* = -.369, *p* = .015) and TOLT highest problem level achieved, (*r* = -.572, *p* = .001). Significant correlations were also observed between the ABS-RC 2 community self sufficiency scale and TOLT total score (*r* = .463, *p* = .002) and TOLT highest problem level achieved (*r* = .502, *p* = .001). ABS-RC 2 personal – social responsibility scale correlated with TOLT total score (*r* = .457, *p* = .002) and TOLT correct at first attempts (*r* = .442, *p* = .003). ABS-RC 2 social adjustment score correlated with TOLT highest problem level achieved (*r* = .458, *p* = .002). ABS-RC 2 personal adjustment score correlated with TOLT total score, (*r* = .472, *p* = .002), TOLT correct at first attempts, (*r* = .456, *p* = .002) and TOLT highest problem level achieved (*r* = .506, *p* = .001).

###### *2.4.1.3.4 Reliability.*

There are no known data attesting to the reliability of the TOLT in an ID sample. Despite this, Willner et al. (2010a) found no significant difference between mean total score on the TOLT in a sample of 40 people with mild to moderate ID and the mean score on the same test used by Ball et al. (2008) in a sample of 78 people with Down’s Syndrome and mild to moderate ID.

### 2.4.2 Everyday symptoms of the dysexecutive syndrome.

A measure of everyday symptoms of executive dysfunction was required in order to assess how well the MET-ID would correlate with these symptoms.

#### *2.4.2.1 The Dysexecutive Questionnaire.*

##### *2.4.2.1.1 Description.*

Burgess et al. (1996) developed the DEX as a means of assessing a range of symptoms associated with the dysexecutive syndrome. The DEX contains 20 items each responded to on a five point Likert type scale (0-4), asking about the degree to which an individual conforms to the specified items (ranging from *never* to *very often*). The DEX comes in both self and independent rater forms. The independent rater form (DEX-IR) was used here. Whilst, the DEX is based on symptoms of the dysexecutive syndrome, some have regarded it as a measure of the SAS (Chan et al., 2008). The characteristics of the dysexecutive syndrome as measured by the DEX questionnaire are outlined in Table 1.1 in the introduction section.

##### *2.4.2.1.2 Administration and scoring.*

A member of staff at the participant's day centre was asked to complete the DEX-IR. The DEX-IR gives a score out of 80, with higher scores equating to more executive dysfunction.

*2.4.2.1.3 Validity.*

Burgess et al. (1998) found that DEX-IR ratings were significantly higher in 92 patients with acquired brain injury or dementia compared to 217 neurologically healthy controls ($p < .001$). Burgess et al. also examined the relationship between the DEX questionnaire and several measures of executive function in the clinical sample. The measures included the Cognitive Estimates Test (Shallice & Evans, 1978), a Verbal Fluency Test (Benton, 1968), the MWCST, the SET and the Trail Making Test (Armitage, 1946). They found significant correlations between DEX-IR scores and performance on all the above tests (apart from the Cognitive Estimates Test) between $r = .29$ and .40 (scores reflected where appropriate). Prior to this Wilson et al. (1996) demonstrated significant correlations (between $r = -.31$ and -.46) across all BADS subtests and DEX-IR scores (all $p < .01$), where low BADS scores mean more impairment whereas high DEX-IR scores mean more impairment. They also found that BADS total profile score showed a significant correlation with DEX-IR score ($r = -.62$, $p < .001$).

The only known study incorporating the DEX-IR in an ID sample is by Masson et al. (2010). As stated above, they found significant correlations between DEX-IR scores and the three TOLT outcome measures (between $r = -.369$ and -.572, all $p < .015$).

#### *2.4.2.1.4 Reliability.*

Bennett et al. (2005) measured the internal consistency of the DEX in four different groups of people. Here, a neuropsychologist, an occupational therapist and family members of participants with brain injuries rated 64, 45 and 42 brain injured participants respectively on the DEX-IR. Fifty-five brain injured participants also rated themselves on the self report version of the DEX. Cronbach's alpha coefficient scores across all groups were higher than .91. Inter-rater agreement on DEX-IR ratings between the neuropsychologist and the occupational therapist when each assessing 30 brain injured participants was high, $r$ = .79, $p$ < .01. In contrast, the correlation between the DEX-IR ratings of neuropsychologists and the family members of 42 brain inured participants was lower $r$ = .42, $p$ < .01, as was the correlation between the DEX-IR ratings of occupational therapists and the family members of 30 brain inured participants, $r$ = .45, $p$ < .05. Whilst the correlation between the self report version of the DEX, as completed by 42 brain injured participants correlated significantly with their families ratings of them on the DEX-IR, $r$ = .68, $p$ <.01, this was not the case for the neuropsychologists DEX-IR ratings across 55 brain injured participants, $r$ = .24, *ns,* and the occupational therapists ratings across 40 brain injured participants $r$ = .23, *ns*.

### 2.4.3 Intellectual Functioning.

A measure of intellectual functioning was used in order to establish the presence of ID in the participants.

#### *2.4.3.1 The Wechsler Abbreviated Scale of Intelligence.*

##### *2.4.3.1.1 Description.*

The WASI gives an abbreviated means of obtaining age adjusted Verbal IQ (VIQ), Performance IQ (PIQ) and Full Scale IQ (FSIQ) estimates for those aged 6 to 89. It contains four subtests which are alternate forms of the Vocabulary, Similarities, Block Design and Matrix Reasoning subtests from the Wechsler Adult Intelligence Scale – Third Edition (WAIS-III; Wechsler, 1998). These subtests have been shown to load well onto the measurement of general intelligence (Spearman, 1927), with values of between .72 and .83 (Kaufman & Lichtenberger, 1999).

##### *2.4.3.1.2 Administration and scoring.*

The WASI takes approximately 30 minutes to administer and was delivered according to manual guidelines. It provides standardized scores for each of the subscales, including a FSIQ estimate which can be derived from administering either two or all four (WASI FSIQ-4) of the WASI subtests. WASI FSIQ-4 was used here, with a score of below 70 used to classify the presence of ID.

*2.4.3.1.3 Validity.*

In Children, Wechsler (1999) reports that the WASI Vocabulary, Similarities and Block Design subtests have demonstrated correlations (between $r$ = .69 and .74) with their alternate forms from the Wechsler Intelligence Scale for Children – Third Edition (WISC-III; 1992). Because the WISC-III does not have a Matrix Reasoning subtest, this comparison could not be calculated. Here, WASI FSIQ-4 also correlated well with WISC-III FSIQ ($r$ = .81). In adults, Wechsler (1999) reports that the WASI Vocabulary, Similarities, Matrix Reasoning and Block Design subtests have demonstrated correlations (between $r$ = .66 and .88) with their alternate forms from the WAIS-III. WASI FSIQ-4 also correlated equally well with WAIS-III FSIQ ($r$ = .92).

The WASI was also administered to 119 individuals with intellectual disabilities. This included 57 individuals with mild mental intellectual disabilities, 30 individuals with moderate intellectual disabilities and 32 individuals with Down's Syndrome. The mild intellectual disability group obtained a mean WASI FSIQ-4 score of 63.28 (*SD* = 7.82), the moderate intellectual disability group obtained a mean WASI FSIQ-4 score of 56.73 (*SD* = 3.74) and the Down's Syndrome group obtained a mean WASI FSIQ-4 score of 57.88 (*SD* = 4.75). Ninety seven percent of the overall sample obtained WASI FSIQ-4 scores of 75 or lower. This dropped to 87% of the sample when WASI FSIQ-4 score was set at 70.

#### *2.4.3.1.4 Reliability.*

For the child sample, Wechsler (1999) found the internal consistency across all subtests to be .87 to .92. For the adult sample, this ranged from .92 to .94. For the IQ scores, the average reliability coefficients for the child sample were .93, .94 and .96 for the VIQ, PIQ and FSIQ-4 scales respectively. For the adult sample, the average reliability coefficients were .96, .96, and .98 for the VIQ, PIQ and FSIQ-4 scales respectively.

Wechsler (1999) also reports test re-test reliability in the child sample ranging from .77 to .86 for the subtest scores, and .88 to .93 for the IQ scales. In the adult sample, this was between .79 to .90 for the subtests, and .87 to .92 for the IQ scales. Four raters were also used to assess inter-rater reliability by scoring 60 protocols. Across all ages, inter-rater reliability of the subtests and IQ scales, ranged from .77 to .92.

### 2.4.4 Receptive vocabulary.

A measure of receptive vocabulary was employed so the participants' ability to understand the MET-ID instructions could be controlled for in the analysis.

#### ***2.4.4.1 The British Picture Vocabulary Scale – Third Edition.***

##### *2.4.4.1.1 Description.*

The British Picture Vocabulary Scale – third edition (BPVS-III; Dunn, Dunn, Styles, & Sewell, 2009) is a measure of receptive (or hearing) vocabulary and is regarded as an indicator of the extent of English vocabulary acquisition (Dunn et al., 2009). It is recommended by the authors for use in children with ID and special educational needs. The BPVS has been used in previous investigations of executive function in adults with ID as a measure of non executive ability (Ball et al., 2008; Willner et al., 2010a).

##### *2.4.4.1.2 Administration and scoring.*

The BPVS-III takes approximately 10 minutes to administer. It requires the examiner to show the participant a page containing four pictures and read out a word. The participant is required to point to the picture that best relates to the word. Outcome measures include raw score, age standardized score and age equivalent score.

##### *2.4.4.1.3 Validity.*

Dunn et al. (2009) found that the BPVS-III correlated with the Cognitive Ability Tests (CAT; Lohman & Hagen, 2001) Verbal Battery ($r$ = .72), as well as CAT composite score ($r$ = .61) in children, where higher scores on both tests mean better performance (significance values not given). Dunn et al. (2009) compared 27 students with a statement of special educational needs with standardised scores of school students of a similar age without special educational needs. They found that the special educational needs group returned a mean BPVS-III score of 88.5 (*SE* = 2.6) which was significantly lower than the normative mean of 100 ($p$ < .05).

##### *2.4.4.1.4 Reliability.*

No known research exists investigating the reliability of the BPVS-III in an ID sample. Therefore, its reliability is assumed based on the reliability of its predecessor (the BPVS-II) in an ID sample. Here, Glenn and Cunningham (2005) found a median Cronbach's alpha of .93 and median split half reliability of .85.

### **2.4.5 Reading ability.**

A measure of reading ability was employed so the participants' ability to read the written instructions of the MET-ID instructions could be controlled for in the analysis.

#### ***2.4.5.1 Word Reading subtest of the Wechsler Individual Attainment Test – Second UK Edition.***

##### *2.4.5.1.1 Description.*

The Word Reading subtest (WR) of the Wechsler Individual Attainment Test – Second UK Edition (WIAT-II; Wechsler, 2005) assesses the ability of the respondent‘s early reading (phonological awareness), word recognition and word decoding skills by asking them to read aloud lists of letters and words accurately. The test provides normative data from children aged six to adults.

##### *2.4.5.1.2 Administration and scoring.*

Each participant was started on item 48 where the respondent is required to read out whole words. The reversal rule was applied where participants scoring zero on any of the first three items were administered the preceding items in reverse order. Participants obtaining full marks on the first three items proceeded until the discontinue criteria of seven incorrect responses was met.

##### *2.4.5.1.3 Validity.*

Wechsler (2005) administered the WIAT-II to 39 people aged 7 to 14 who had a FSIQ < 70 and were diagnosed with ID. The mean standardized score for the WR subtest in this group was 58.14 (*SD* = 10.42). Eighty five %

of this group obtained WR scaled scores ≤ 70. A matched control group of participants with IQs > 70 obtained a mean standardized score of 95.68 (*SD* = 14.47). Ten percent of the matched control group obtained WR scaled scores ≤ 70. The difference between the mean WR subtest standardized scores across the both groups was statistically significant ($t$ = 12.36, $p$ < .01).

##### *2.4.5.1.4 Reliability.*

Average split half reliability coefficient across ages 6 to 19 on the WR was .97. Across all ages, average test retest reliability of the WR was .98. Inter-rater reliability of all WIAT-II subtests has been found to range from .94 to .98, across all ages, with an overall reliability of .94.

## 2.5 Procedure

For those with ID, each participant was initially approached by a member of staff from their day centre who was well known to them. The member of staff assessed the participant's interest in taking part by explaining the information sheet to them. If an interest was expressed, the CI subsequently approached the participant to confirm their interest before obtaining informed consent and demographic data. Thereafter, each participant was asked to undertake the WASI, BPVS and the WR subtest from the WIAT-II in the first meeting. This took approximately one hour. The CI then arranged another time to meet with the participant within the following week, where the TOLT, SPT and MET-ID were administered.

Again, this second appointment lasted approximately one hour. Following completion of the second session the participant was thanked for their time.

For staff asked to complete the DEX-IR, the day centre manager briefed the staff about the nature of the research and asked for volunteers who would be willing to give ratings. If an interest was expressed, the CI approached each member of staff individually and went over the information sheet, informed consent form and demographic data sheet. Staff participants were then asked to provide DEX-IR ratings for each ID participant until data from the required number was obtained.

## 2.6 Plan of Analysis

Each hypothesis is outlined below, followed by the planned analysis.

1) The DEX-IR will correlate more strongly with the MET-ID scales and the SPT compared to the TOLT as the MET-ID and SPT are designed to be more predictive of the "real life" symptoms of SAS impairment.

2) There will be a stronger correlation between the MET-ID scales and the SPT compared to the correlation between the MET-ID and the TOLT as the MET-ID and SPT are designed to tap a subset of the same cognitive components required in "real life" tasks.

The correlations between each measure were calculated. Meng, Rosenthal and Rubins's (1992) test for comparing correlated correlations within a single sample was then used to assess if there were any significant differences between the magnitudes of these correlations.

# Chapter Three: Results

## 3.1 Overview

This section outlines the results of the study. Initially, descriptive information about the sample used and data collected are presented. The hypotheses are then tested, followed by supplementary analysis examining the influence of receptive vocabulary and reading ability on MET-ID performance. It is concluded that none of the research hypotheses was supported. Equally, the receptive vocabulary, reading ability and general verbal abilities of the participants were strong predictors of MET-ID performance. However, even when receptive vocabulary, reading ability and general verbal abilities were partialled out, MET-ID task completions continued to correlate significantly with performance on the TOLT and its subscales, suggesting that the MET-ID has a robust relationship with this measure.

## 3.2 Descriptive information about sample

Only data from the 40 non-pilot participants were used in the final analysis. Participants were recruited from four different day centres. For the 40 non pilot participants, four (10%) were from day centre A, 16 (40%) were from day centre B, 13 (32.5%) were from day centre C and seven (17.5%) were from day centre D. A Kruskal-Wallis test revealed no significant differences in mean MET-ID task attempts ($H$ (3) = 1.878, $p$ = .598), task completions ($H$ (3) = .224, $p$ = .974) and rule breaks, ($H$ (3) = 2.356, $p$ = .887) across the four day centres.

### 3.3 Descriptive data analysis

Twenty five percent of the MET-ID scripts were scored by both the author and the author's primary research supervisor. The intraclass correlation coefficient for task attempts was .885, for task completions .986 and for rule breaks .840. For all analyses below, the scoring undertaken by the principal investigator was used.

Internal consistency was calculated for each of the MET-ID subscales. Cronbach's alpha coefficient was .838 for task attempts and 611 for task completions. However, for rule breaks, Cronbach's alpha coefficient was -.584. Accordingly, the correlation matrix comparing the frequency with which the respective rule breaks were adhered to or broken revealed a high degree of negative relationships. For example, the rules "Do all the tasks in any order" and "Don't speak to Tom unless it's to tell him you have finished" were broken by 77% and 72.5% of the participants respectively. In contrast, the rules "Don't walk into any staff offices," "Don't walk back into a room you've already been in" and "Don't leave the day centre" were only broken by 12.5%, 17.5% and 5% of the sample respectively. This suggests that the rules broken most frequently may have been difficult for the participants to avoid breaking (e.g., it was difficult for the participants to complete all the tasks and not speak to the examiner), whereas the rules that were broken less frequently may have been easy to adhere to (e.g., it was easy for the participants to avoid walking into any staff offices, avoid walking back into a room and not leave the day centre).

Another difficulty with the rule breaks measure was that, as outlined above, the very act of attempting the tasks makes the participant more likely to break rules compared to participants who fail to attempt any tasks. Thus, a participant who attempted many tasks may actually receive a worse error score profile than a participant who didn't attempt as many as attempting more tasks could mean they could potentially break more rules. Accordingly, a positive correlation would be expected between task attempts and rule breaks, and task completions and rule breaks. As can be seen in Table 3.2 this prediction occurred, where rule breaks were significantly positively related to task attempts and task completions. This further questions the validity of the rule breaks on this version of the MET-ID measure as it appears highly task dependent.

In addition, the very act of being able to adhere to the rules is dependent on the participants' ability to be aware of them. Whilst the participants were given one minute to help them learn the rules when the MET-ID was being explained, responses indicated that when asked to recall them, a mean of 2.02 out of six were spontaneously recalled correctly at first attempt. When the participant's were given prompts to help them remember each individual rule a mean of 1.50 additional rules were recalled at second attempt. Thus, after both spontaneous and prompted recall, an average of 3.52 rules were remembered by the participants prior to attempting the MET-ID. Therefore, overall it is plausible that participants had great difficulty actually holding these rules in mind when completing the MET-ID (even though they had the rules written down in front of them on the MET-ID exercise sheet when completing the task).

Accordingly, based on its poor internal consistency, the finding that rule breaks were highly task dependent and the poor memory participants had for the rules, the rule breaks scale on this particular version of the MET appears to have poor reliability and validity. However, for illustration, it will still be used in the analysis below.

For the remainder of the data, SPSS boxplots revealed that only WASI and DEX-IR scores had outliers. These included two participants with IQs of 69 and 70 respectively and one participant with a DEX-IR score of 67. Analysis of the dataset with these participants removed produced no meaningful influence on the results. Therefore, the data reported below has these outliers included. Table 3.1 shows descriptive statistics for the MET-ID scales, SPT, TOLT, DEX-IR, WASI, BPVS and WR.

Table 3.1

*Descriptive statistics for the measures*

| Measure | Maximum score | No at minimum (%) | No at maximum (%) | Range of scores | Mean (SD) | Median |
|---|---|---|---|---|---|---|
| Task attempts | 6 | 4 (10) | 10 (25) | 0 – 6 | 3.9 (2.1) | 5 |
| Rule breaks | 6 | 0 (0) | 0 (0) | 1 – 4 | 2.2 (0.8) | 2 |
| Task completions | 6 | 0 (0) | 4 (10) | 1 – 6 | 3.3 (1.5) | 3 |
| SPT | 16 | 0 (0) | 0 (0) | 1 – 13 | 6.3 (3.0) | 6 |
| TOLT | 18 | 0 (0) | 0 (0) | 2 – 17 | 10.5 (4.8) | 10.5 |
| DEX-IR | 80 | 0 (0) | 0 (0) | 6 – 67 | 30.6 (14.5) | 29 |
| WASI | 130 | 0 (0) | 0 (0) | 52 – 70 | 58 (4.54) | 56.5 |
| BPVS | 169 | 0 (0) | 0 (0) | 19 – 159 | 102 (31.2) | 99.5 |
| WR | 131 | 0 (0) | 0 (0) | 3 – 124 | 59.8 (36.0) | 54.5 |

*Note.* BPVS = British Picture Vocabulary Raw Score; DEX-IR = Dysexecutive Questionnaire Independent Rater total score; SPT = Six Parts Test total score; TOLT = Tower of London Test total score; WASI = Wechsler Abbreviated Scale of Intelligence Full Scale IQ; WR = Wechsler Individual Attainment Test word reading subtest total score.

The degree of skewness and kurtosis for each measure is outlined in Appendix P. Kolmogorov-Smirnov and Shapiro-Wilk tests for each measure are given in Appendix R. In sum, the MET-ID scales, WASI FSIQ and TOLT total scores were not normally distributed. In contrast, BPVS raw score, WR raw score, SPT total score and DEX-IR total score were normally distributed. As the MET-ID subscales were the variables of interest, non parametric Spearman’s correlations were used to compare the data. A correlation matrix between the MET-ID subscales, SPT, TOLT, DEX-IR, VIQ, PIQ, FSIQ, BPVS and WR scores is presented in Table 3.2. High DEX-IR scores and MET-ID rule breaks equate to worse executive function whereas high MET-ID task attempts, MET-ID task completions, TOLT and SPT scores equate to better executive function.

Table 3.2

*Summary of Spearman's rho correlations between the measures used*

| Measure | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Task attempts | - | | | | | | | | | | |
| 2. Rule breaks | .438** | - | | | | | | | | | |
| 3. Task completions | -.732** | -.431** | - | | | | | | | | |
| 4. SPT | .214 | -.150 | .267* | - | | | | | | | |
| 5. TOLT | .222 | -.027 | .368* | .080 | - | | | | | | |
| 6. DEX-IR | .050 | .044 | .124 | .142 | -.016 | - | | | | | |
| 7. VIQ | .328* | .238 | .354* | .291 | -.116 | -.099 | - | | | | |
| 8. PIQ | .000 | -.017 | .067 | .088 | .379* | -.306 | .206 | - | | | |
| 9. FSIQ | .190 | .061 | .243 | .282 | .335* | -.344* | .481** | .887** | - | | |
| 10. BPVS | .392* | .102 | .432** | .258 | .219 | -.142 | .451** | .358* | .465** | - | |
| 11. WR | .346* | .087 | .257 | .231 | -.046 | -.285 | .576** | .306 | .492** | .351* | - |

*Note.* Significance values above the dotted line are one tailed. Significance values below the dotted line are two tailed. BPVS = British Picture Vocabulary Scale Score; DEX-IR = Dysexecutive Questionnaire Independent Rater total score; FSIQ = Wechsler Abbreviated Scale of Intelligence Full Scale IQ; PIQ = Wechsler Abbreviated Scale of Intelligence Performance IQ; SPT = Six Parts Test total score; TOLT = Tower of London Test total score; VIQ = Wechsler Abbreviated Scale of Intelligence Verbal IQ; WR = Wechsler Individual Attainment Test – Second Edition Word Reading subtest total score.

* $p < .05$

** $p < .01$

## 3.4 Main research questions

The hypotheses are now reviewed and tested in turn.

1) The DEX-IR will correlate more strongly with the MET-ID scales and the SPT compared to the TOLT as the MET-ID and SPT are designed to be more predictive of the “real life” symptoms of SAS impairment.

As can be seen in Table 3.2, the correlations between the DEX-IR and measures of executive function were all small and non-significant. When the subscales of the DEX-IR were correlated with these measures, there were again no significant relationships. Accordingly, testing hypothesis one is unnecessary because the hypothesis was not supported.

2) There will be stronger correlations between the MET-ID scales and the SPT compared to the correlation between the MET-ID and the TOLT, as the MET-ID and SPT are designed to tap a subset of the same cognitive components required in “real life” tasks.

As can be seen in Table 3.2, only MET-ID task completions correlated significantly with the SPT and the TOLT and the correlation between task completions and the TOLT was numerically higher than the correlation between task completions and the SPT. Therefore, again testing hypothesis two is unnecessary because the hypothesis was not supported.

## 3.5 Additional data analysis

In order to control for the effects of IQ (WASI VIQ, PIQ and FSIQ), receptive vocabulary (BPVS score) and reading ability (WR score) on MET-ID performance, the correlation matrix was analysed to assess whether any of these variables was related to MET-ID performance. As can be seen in Table 3.2, VIQ was found to be significantly correlated with MET-ID task attempts and task completions. Also, receptive vocabulary (BPVS score) was found to be significantly correlated with MET-ID task attempts and task completions. Equally, reading ability (WR score) was found to be significantly correlated with task attempts.

Accordingly, Table 3.2 was re-consulted to assess if the IQ, BPVS and/or WR scores that correlated with the MET-ID scales, also correlated with the other measures of executive function. This was to test if these third variables (e.g., IQ, BPVS and/or WR scores) accounted for some of the variance in the correlations between the measures of executive function. Only if these third variables correlated with both the MET-ID scales and the SPT, DEX-IR and/or TOLT would they be included in this analysis, as it would be likely that the third variables were influencing performance on both the measures and therefore would account for some of the variance in the correlations between the MET-ID and the executive function tests. Accordingly, when one tailed significance tests were applied to the correlations below the dotted line in Table 3.2, BPVS score was found to be significantly related to MET-ID task completions ($r_s$ = .432, $p$ = .003, one tailed) and SPT score ($r_s$ = .258, $p$ = .038, one tailed). On top of this, VIQ was also found to correlate with both MET-ID task completions ($r_s$ = .354, $p$ =

.012, one tailed) and SPT score ($r_s$ = .291, $p$ = .034, one tailed). Therefore, a partial correlation was conducted to control for the effects of BPVS score and VIQ on the relationship between MET-ID task completions and the SPT. As SPSS 19 does not allow partial correlations to be carried out on a Spearman's correlation, the data were manually ranked and a Pearson's correlation (on which SPSS 19 does allow for the partialling out of the variables) was carried out on the ranked data. When BPVS score and VIQ were simultaneously partialled out, the relationship between MET-ID task completions and the SPT was no longer significant ($r_s$ = .134, $p$ = .211, one tailed). Thus, when the effects of BPVS score and VIQ were controlled for, the relationship between MET-ID task completions and SPT performance disappeared, suggesting that receptive vocabulary and general verbal abilities were having a strong influence on the participants' ability to perform well on both these scales.

Whilst, BPVS score and VIQ did not correlate significantly with the TOLT total score, an additional analysis was conducted to assess the influence of these variables (as well and WR score) on the significant positive correlation between MET-ID task completions and the TOLT total score. This was in order to test the degree to which receptive vocabulary, reading ability and general verbal skills were influencing the robustness of this relationship. When BPVS, WR score and VIQ scores were simultaneously partialled out, the relationship between MET-ID task completions and TOLT total score actually strengthened ($r_s$ = .379, $p$ = .010, one tailed). In exploring the relationship between the MET-ID and the subscales of the TOLT, MET-ID task completions correlated significantly with

TOLT highest problem level achieved, ($r_s$ = .328, $p$ = .019, one tailed) and TOLT correct first attempts ($r_s$ = -.393, $p$ = .006, one tailed). When BPVS, WR and VIQ scores were simultaneously partialled out from these correlations, the relationships again strengthened. Here, Spearman's correlations indicated that MET-ID task completions correlated significantly with TOLT highest problem level achieved ($r_s$ = .351, $p$ = .017, one tailed) and TOLT correct first attempts ($r_s$ = .417, $p$ = .005, one tailed). This suggests that the relationship between MET-ID task completions and the TOLT is robust.

An alternate area of interest was the explanation as to why some participants were able to initiate tasks whereas others were not. As outlined above, this could have plausibly been down to failures in task comprehension or, alternately, an executive deficit that prevented participants from being able to formulate plans to initiate any of the tasks. To investigate this, it was hypothesised that lower scores in receptive vocabulary and reading ability (BPVS and WR scores respectively) in those who were unable to initiate any tasks would support a failure in task comprehension. This was based on the reasoning that poor receptive vocabulary would have meant that certain participants were unable to understand the task instructions when explained to them by the examiner or when vocalised by the auditory equipment, and poor reading ability would have meant that certain participants would have been unable to read the tasks/rules on the MET-ID exercise sheet adequately. In contrast, lower scores on the measures of executive function in those who were unable to initiate any of the tasks would support the notion of these participants suffering from an executive deficit, thus accounting for their poorer performance on the MET-ID. Accordingly, participants were divided into two groups based on whether they were able to initiate tasks or not. Participants were allocated to the “non initiators” group if they (a) failed to initiate any tasks at all or (b) only initiated the “tell the examiner when you have finished” task without attempting any other tasks. Participants were ascribed to the “initiators” group if they initiated two or more tasks.

Out of the final 40 participants, a total of nine comprised the non initiators group. The remaining 31 participants comprised the initiator group. One tailed tests were used to assess the aforementioned hypotheses. A Mann-Whitney U test showed that BPVS score in non initiators (*Mdn* = 91.00) was lower than initiators (*Mdn* = 115.00) and that this difference was significant ($U$ = 82, $z$ = -1.863, $p$ = .031, one tailed). Non initiators (*Mdn* = 26.00) were also significantly lower than initiators (*Mdn* = 75.00) on WR score ($U$ = 78, $z$ = -1.992, $p$ = .023, one tailed). There were however no significant differences in TOLT, SPT, DEX-IR, WASI FSIQ and PIQ scores between the two groups. Despite this, the difference between VIQ was significantly lower in non initiators (*Mdn* = 55) compared to initiators (*Mdn* = 55, $U$ = 96, $z$ = -1.694, $p$ = .041, one tailed). Overall therefore, it appears that deficits in non responders reading comprehension, receptive vocabulary and VIQ compromised their capability to understand the requirements of the MET-ID and initiate the tasks over and above any executive deficit.

### 3.6 Summary of data analysis

The hypotheses of the study were not supported. The DEX-IR failed to correlate more strongly with the MET-ID scales and the SPT in comparison to the TOLT. Equally, the MET-ID scales failed to correlate more strongly with the SPT compared with the TOLT. Nevertheless, MET-ID task completions were significantly correlated with performance on both the SPT and the TOLT, where better SPT and TOLT performance meant more MET-ID task completions. Despite this, both MET-ID task completions and SPT performance correlated significantly with BPVS score and VIQ. When BPVS score and VIQ were partialled out, the correlation between the MET-ID and the SPT disappeared suggesting that receptive vocabulary and general verbal abilities were influencing the relationship between MET-ID task completions and the SPT. However, partialling out BPVS score, WR score and VIQ had no effect on the relationship between MET-ID task completions and the TOLT and its subscales suggesting that receptive vocabulary, reading ability and general verbal abilities were not influencing the relationship between MET-ID task completions and the TOLT. Nevertheless, when the scores of those who were unable to initiate any MET-ID tasks were analysed, the data showed that these participants had poorer BPVS, WR and VIQ scores compared to those who were able to initiate the tasks. Thus, overall, the MET-ID has a robust executive component as evidenced by its relationship with the TOLT. However, perhaps unsurprisingly it also loads heavily onto receptive vocabulary, reading ability and VI

# Chapter Four: Discussion

## 4.1 Overview

This chapter discusses the findings of the study. Initially, the research hypotheses are evaluated in the context of the wider literature and additional findings are discussed. The strengths and weaknesses of the methodology are then reviewed with specific consideration given to the design, measures, sample and procedure adopted by the study. Recommendations for how the research needs to be improved and developed are then given. The theoretical implications of the results are then discussed, making reference to the degree to which the findings can be conceptualised within the SAS and/or Duncan's theory of goal neglect. Potential for the adaptation of specific clinical interventions that stem from these theories for people with ID is then considered. Finally, the section concludes that the MET-ID needs further development and adaptation to make it better suited to an ID population. Nevertheless, the findings of this preliminary study offer directions for the future development of measures of EF for people with ID, as well as useful theoretical insights for potential interventions that could be adapted to help improve executive abilities in people with ID.

## 4.2 Evaluation of research hypotheses

This section considers each research hypothesis in turn and summarises the results before relating them to the existing literature.

1) The DEX-IR will correlate more strongly with the MET-ID scales and the SPT compared to the TOLT as the MET-ID and SPT are designed to be more predictive of the “real life” symptoms of SAS impairment.

There was no support for hypothesis one. The correlations between the DEX-IR and all other measures of EF were small and non significant. This is out of context with the wider literature in both ID and non ID samples where the DEX-IR has been shown to correlate significantly with the MET, SPT and TOLT (Chan, 2001; Dawson et al., 2009; Emsile et al., 2003; Knight et al., 2002; Masson et al., 2010). This finding therefore limits the claims the MET-ID can make towards being an ecologically sensitive measure of the SAS.

Nevertheless, other research has shown that the DEX-IR does not always correlate well with measures of EF. For example, in a brain injured sample Norris and Tate (2000) found that the DEX-IR shared small to medium positive correlations (the opposite direction to that expected) with four subtests from the BADS (including the SET) and BADS total profile score. The potential reasons behind why the DEX-IR may not consistently correlate with measures of EF, as found in this study, are discussed in section 4.4.2.4 below.

2) There will be a stronger correlation between the MET-ID scales and the SPT compared to the correlation between the MET-ID and the

TOLT as the MET-ID and SPT are designed to tap a subset of the same cognitive components required in “real life” tasks.

There was no support for hypothesis two. Whilst the correlations between the MET-ID scales and the measures of EF all fell in the anticipated directions, the correlations between the MET-ID scales and the SPT were not significantly stronger than the correlations between the MET-ID scales and the TOLT. Moreover, it was only MET-ID task completions that shared significant correlations with both the SPT and the TOLT, and the correlations between the MET-ID and the TOLT were actually (non significantly) stronger than the correlation between the MET-ID and the SPT. This finding is theoretically disappointing for the MET-ID because, the original versions of the MET and SPT were designed to be sensitive to impairments of the SAS that are not readily picked up by traditional neuropsychological assessments (such as the TOLT) as well as tap a subset of the same cognitive components (Burgess, 2000; Shallice & Burgess, 1991). Therefore, it was expected that the MET-ID should share a stronger correlation with the SPT than the TOLT. The fact that this failed to occur suggests that the MET-ID and SPT may need further adaptation to improve their ability to be ecologically sensitive measures of the SAS in an ID population.

Equally, despite being statistically significant, the correlations between MET-ID task completions and the SPT and TOLT (before controlling for confounding variables) were only small to medium in strength. Nevertheless, Field (2009) emphasises the importance of interpreting such effect sizes within the context of the wider research literature. Indeed, Burgess and

Robertson (2002) outline how correlations between performance on different executive tasks in both neurological and healthy samples are typically low (e.g., Duncan et al., 1997; Emslie et al., 2003; Myakie et al., 2000; Norris & Tate, 2000; Robbins, 1998). Therefore, the strength of the correlations observed here are not out of context with those observed between executive tests in other populations. The implications of these findings are discussed in sections 4.4.2.1 and 4.6 below.

### 4.3 Additional findings

Further analysis was used to control for the effect of receptive vocabulary and reading comprehension on MET-ID performance. Here, reading ability was significantly positively correlated with the number of MET-ID tasks attempted. Equally, receptive vocabulary was significantly positively correlated with the number of MET-ID tasks attempted and tasks completed. It was also found that general verbal abilities (VIQ) were significantly positively correlated with MET-ID tasks attempted and tasks completed.

Furthermore, it was found that both receptive vocabulary and general verbal abilities correlated significantly with both MET-ID task completions and SPT score. When these potentially confounding variables were controlled, the correlation between MET-ID task completions and the SPT weakened to the degree that it was no longer significant. This suggests that these abilities had a strong influence on the relationship between the MET-ID and the SPT. In contrast, whilst no potentially confounding variables correlated with both MET-ID task completions and the TOLT, simultaneously controlling for these variables (e.g., receptive vocabulary, reading ability and general verbal abilities) in the relationship between MET-ID task completions and the TOLT (and it's subscales) did not influence the strength of the relationship. This suggests that the relationship between MET-ID task completions and the TOLT is robust.

Accordingly, further analysis was conducted to assess if there was a difference between participants who were unable to initiate any tasks on the MET-ID compared with those who were. Results showed that those who were unable to initiate any tasks scored lower on receptive vocabulary, reading ability and general verbal abilities compared to both the sample as a whole. Whilst this may be unsurprising, as it has been noted that it would be very difficult to design a measure of executive function that doesn't tap areas such as language (Baddeley et al., 1997), the above findings suggest that MET-ID performance is heavily reliant on receptive vocabulary and reading ability.

## 4.4 Strengths and weaknesses of the methodology

### 4.4.1 Design.

The study used a cross sectional, correlational design using a single sample of participants with ID so that measures of the SAS could be compared with observer ratings of executive function. This design had a number of strengths. Firstly, the correlational approach allowed for potentially confounding variables to be assessed and controlled in the analysis. This allowed firm conclusions to be drawn about the extent to which the MET-ID was measuring executive function over and above reading ability and receptive vocabulary, as well as the influence these variables had on the relationship between the MET-ID scales and the alternate measures of EF.

The use of Meng et al.'s (1992) *z* test for comparing correlated correlations would have also allowed hypotheses to be tested about whether the MET-ID correlated more strongly with one variable over another. This would potentially have allowed firm conclusions to be drawn about the strengths of the correlation coefficients, had the results lent themselves to this analysis.

Moreover, correlating the executive function measures with the DEX-IR allowed an estimate of the degree to which the measures predict the participant's behaviour in real life. This enabled firm conclusions to be drawn about the ecological validity of the measures, which is an increasingly pressing concern in neuropsychological assessment (Burgess et al., 2006).

### 4.4.2 Measures.

To assess executive function, the study employed an adapted version of the MET (MET-ID), the SPT, an adapted version of the TOLT and the DEX-IR. These measures were chosen because they have been specifically developed as tests of the SAS (Chan et al., 2008). However, due to the relative scarcity of measures specifically developed to test the SAS in an ID population, conclusive data on the reliability and validity of these measures in an ID population is largely un-established. Accordingly, these issues are discussed immediately below, along with the respective strengths and weaknesses of the measures.

#### *4.4.2.1 The MET-ID.*

The MET-ID, as developed here, has a number of strengths. Firstly, this was the only known study in which the same version of the MET has been applied across a range of different environments. There proved to be no differences in MET-ID scores across the four different day centres. This is encouraging as a test for the future development and utility of the MET will be its ability to be administered in a consistent way across a range of settings.

Indeed, adapting the MET-ID for day centres was useful because these are easily accessible, safe and convenient environments for psychologists to assess the real world cognitive abilities of people with ID. Accordingly, such assessment can afford psychologists a much richer understanding of how an individual’s cognitive skills operate outside of structured, office based settings and provide useful quantitative and qualitative information to support formulation, intervention and re-assessment.

In many cases, the use of auditory equipment to vocalise the written instructions of the MET-ID also helped to circumvent the difficulties faced by many of those who were unable to read the MET-ID instructions. This is encouraging for the development of the MET-ID as it shows that such equipment can support the ability of people with ID to operate independently in unstructured environments, which is a key part of the MET-ID assessment procedure.

It should also be noted that MET-ID task completions shared a robust relationship with the TOLT and its subscales. This is encouraging for the MET-ID because the TOLT is possibly one of the most widely used measures of EF in people with ID suggesting that it is a highly accessible measure for this population (Adams & Oliver, 2010; Ball et al., 2008; Masson et al., 2010; Willner et al., 2010a). Thus, the fact that the TOLT correlates well with MET-ID task completion suggests that an (in)ability to initiate tasks on the MET-ID may be tapping a specific executive process in people with ID (see sections 4.5 and 4.6 below).

Despite this, as previously mentioned, the fact that the correlation between MET-ID task completions and the SPT disappeared when receptive vocabulary (BPVS) and general verbal abilities (VIQ) were controlled is clinically and theoretically disappointing for the MET-ID. This is because both measures are designed to be ecologically sensitive and tap a subset of the same cognitive components of the SAS (Burgess, 2000; Shallice & Burgess, 1991). Therefore, the lack of correlation between them suggests that they may not be performing these functions in the present sample. Equally, despite the fact that MET-ID task completions correlated well with the TOLT, the TOLT has been criticised for its lack of ecological validity (Burgess et al., 2006; Burgess & Robertson, 2002, see also section 4.4.2.3 below). Therefore, beyond "face" ecological validity, there remain unanswered questions about the extent to which the MET-ID does tap executive skills in the real world, especially in view of the fact that it failed to correlate with the

DEX-IR (notwithstanding the DEX-IR's unestablished validity and reliability in an ID sample, see section 4.4.2.4).

Nevertheless, in terms of reliability, scoring of the written scripts taken from MET-ID performance showed a high degree of inter-rater agreement across the MET-ID scales with the lowest interclass correlation coefficient being .840 for task completions. Equally, internal consistency for task attempts was high (.838). However, it was relatively moderate for task completions (.611). Nevertheless, for rule breaks, internal consistency was poor (-.584). As outlined above, this is likely to be because approximately 75% of the sample broke the rules "Do all the tasks in any order" and "Don't speak to Tom unless it's to tell him you have finished" whereas the rules "Don't walk into any staff offices," "Don't walk back into a room you've already been in" and "Don't leave the day centre" were each only broken by less than 20% of the sample. This discrepancy between high frequency and low frequency rule breaks lead to a number of negative correlations across the frequency with which the rule breaks were committed (e.g., some rule breaks were committed a lot, whereas others were rarely committed at all), hence the poor internal consistency.

Another problem with the rule breaks measure was that the participants had a poor memory for the rules after they had been specifically asked to learn and remember them. Thus, if the participants were not aware of, easily forgot or were unable to read the rules when completing the MET-ID, it is unsurprising that they may be liable to breaking some of them (even though they did have the rules written in front of them and could use the pen to vocalise them when they were executing the task). This is unfortunate for the MET-ID as the rule breaks scale has been shown to correlate with various measures of executive function (Alderman et al., 2003; Knight et al., 2002) as well as being predicative of distinctive dysexecutive symptoms (Alderman et al., 2003). This suggests that rule breaks can be a highly clinically useful scale and it is unfortunate that an acceptable degree of performance was not captured here. Thus, the rules of the MET-ID need further adaptation in order to make rule breaks a more clinically useful scale (see section 4.5 below).

#### *4.4.2.2 The SPT.*

The SPT was chosen for this study because of the theoretical relationship between the SET (the adult version of the SPT) and the MET (Burgess, 2000; Shallice and Burgess, 1991). The SPT was used (rather than the SET) because it is taken from the BADS-C and was therefore considered to be a more developmentally appropriate measure. Psychometrically, the SPT had some strength in this sample. The scores were normally distributed and there were no floor or ceiling effects.

However, whilst the SPT was designed to be used with children, its validity in an ID sample is un-established. Indeed, the only known investigation examining the SPT in an ID sample was by Willner et al. (2010a), as part of a wider investigation into the CEFA and the BADS-C. As outlined above, Willner and colleagues found that the SPT shared a small but significant correlation with the TOLT. However, in the present sample, no meaningful correlation was found between the SPT and the TOLT. Moreover, whilst Emslie et al. (2003) found the SPT to have a medium, significant correlation with the children's version of the DEX, there was no meaningful relationship between the SPT and the DEX-IR in this sample. Equally, qualitative observation of the participant's performance suggested that this task was poorly understood by the participants. Although every effort was made to ensure that the participants understood the requirements of the task, explaining the SPT is a long verbal procedure and it was subjectively felt that the majority of participants failed to grasp what they were being asked to do. For example, only one participant actually managed to complete the whole task without breaking the rule, despite all participants being made explicitly aware of it. Accordingly, the SPT is likely to need further adaptation to be optimally accessible for this client group.

#### ***4.4.2.3 The TOLT.***

The TOLT was chosen for this study because of its theoretical underpinnings and popularity within the ID literature (Adams & Oliver, 2010; Ball et al., 2008; Masson et al., 2010; Willner et al., 2010a). In the present study, TOLT scores were not normally distributed. However, no floor or ceiling effects were apparent. Equally, as stated, the TOLT and its subscales shared a robust relationship with MET-ID task completions. This is surprising because such relationships failed to transpire when the MET has been compared with the TOLT in alternate populations (Knight et al., 2002). In contrast, the TOLT failed to correlate with the DEX-IR in this study. This runs contrary to alternate work in which the TOLT has shown a robust relationship with the DEX-IR (Masson et al., 2010).

Subjective observation of TOLT performance suggests that most (if not all) participants were able to engage with and understand the task. The administration procedure outlined by Masson et al. (2010) allowed the participants to complete the task in a way that is scaffolded by the examiner (e.g., the participant works through progressive degrees of difficulty and the examiner provides immediate verbal feedback whenever a rule is broken). Whilst this makes it highly accessible for an ID population, very few (if any) real life situations pose executive demands that progressively increase in difficulty and have immediate environmental feedback. Therefore, the face ecological validity of the TOLT continues to be in question (Burgess et al., 2006; Burgess & Robertson, 2002).

#### *4.4.2.4 The DEX-IR.*

The DEX-IR was used in this study to assess how representative the executive measures were of everyday symptoms of executive dysfunction. None of the measures of the SAS correlated meaningfully with the DEX-IR. Accordingly, several potential reasons could explain this. Firstly, the DEX-IR may not be a valid measure in an ID sample as it was originally designed for neurological samples. Otherwise, the study used employees of the day centres as informants on DEX-IR. Whilst these people were expected to have a good knowledge of and familiarity with the participant, it is very difficult to ascertain whether people completing the DEX-IR are doing it with the same degree of awareness and understanding of the participants' behaviours (Chaytor et al., 2006; Norris and Tate, 2000). Indeed, the inter-rater-reliability of the DEX-IR is likely to vary depending on the person doing the rating. For example, Bennett et al. (2005) found that the DEX-IR ratings of neuropsychologists and occupational therapists in assessing brain injured participants correlated significantly with BADS total score, whereas there were no significant correlations between the DEX-IR ratings of the brain injured participant's family members and the brain injured participant's self report DEX ratings with BADS total score. Bennett et al. therefore suggested that the DEX-IR can only be used confidently as a screening instrument to identify executive dysfunction if it is completed by a professional trained to be sensitive to the cognitive and behavioural symptoms of the dysexecutive syndrome. It is unlikely that the people completing the DEX-IR in this study had a firm grasp of this. Moreover, Bennett et al. also point out that the DEX-IR does not identify time frame (e.g., behaviours shown over the last week or last year for example) nor the frequency with which the behaviours are shown (e.g., once per day or once per year etc). This can again distort the ratings and possibly explain why they failed to correlate with the measures used in this study. Accordingly, alternate measures of real life abilities may be more appropriate (see section 4.5 below).

### 4.4.3 Sample.

The sample size was consistent with other investigations of executive function in people with ID (Dymond et al., 2010; Masson et al., 2010; Willner et al., 2010a). Ethical issues in recruitment associated with acquiescence were overcome by asking a key worker to go over the study information sheet with the participants prior to their formal contact with the researcher. Equally, day centre managers were asked to put forward participants who were judged to be able to give informed consent, thus further reducing the potential for acquiescence to occur. Collection of the data from day centres also allowed for informant report data on the DEX-IR to be readily collected. Use of the WASI also enabled precise screening to be carried out to ensure that participants were in the mild ID range (FSIQ < 50 and > 70). Equally, the sample excluded potential participants with dementia, schizophrenia and/or brain injury, as these have been shown to impair the SAS independent of ID (Andrés & Van der Linden, 2001; Marczewski et al., 2001; Perry & Hodges, 1999).

Nevertheless, a potential design problem arising from the sampling of this study was that participants would have been attending day centres on the basis that they are likely to struggle with many of the cognitive skills required in everyday tasks, thus necessitating involvement with day services. This may therefore have meant that the sample included participants who were at the more limited scale of the mild ID range and overlooked participants, who for example, have the cognitive skills that do not require involvement with day services.

#### 4.4.4 Procedure.

A potential problem with the recruitment procedure was that participants were put forward for the study by recommendation of their day centre manager on the basis that they would be able to give informed consent. However, participants who can engage in the decision making processes required to give informed consent may actually demonstrate better EF (see Willner, Bailey, Parry & Dymond, 2010b). Therefore, participants put forward for the study may have been at the higher range of executive functioning, albeit within this potentially more globally cognitive impaired day services sample. Taken together, these issues suggest that caution be exercised in generalising the findings.

### 4.5 How can the research be improved and developed

There are a number of ways in which the MET-ID reported here needs to be improved and developed. Initial pilot testing was carried out to ensure that the MET-ID could capture an appropriate range of performance for an ID sample. However, due to the time constraints involved in this piece of work, the measure finally settled on here was only modified once from its original conception. Thus, it is likely that several further modifications would have benefitted the measure used here.

For example, based on its poor internal consistency, the finding that rule breaks were highly task dependent and the poor memory participants had for the rules, the rule breaks scale failed to capture an appropriate range of errors for it to be a clinically meaningful scale. Here, it was not clear that the participants had an obvious memory for or awareness of the rules (even though they did have the rules in front of them when they were carrying out the MET-ID) as nearly all participants broke the rules of speaking to the examiner and not attempting all the tasks. Accordingly, having to be mindful of six rules may have loaded too heavily on working memory demands so the participants found it difficult to bear them in mind or check them whilst completing the tasks. Indeed, such practice (doing tasks whilst checking rules) may itself present a “dual task” scenario (Baddeley et al., 1997, Della Sala et al., 2010) or tap “switching” or “set shifting” skills that load onto executive abilities and may be problematic for people with ID (Ball et al. 2008; Danielson et al., 2010; Kittler et al., 2008; Lanfranchi et al., 2011). Therefore, it may be useful to reduce the rules of the MET-ID to ensure that they are not overloading the participant’s working memory or ability to shift set in order to attain an appropriate range of performance on this scale. Here, it might be useful to remove the “Don’t walk inside the reception”, “Don’t walk inside any staff offices” and “Don’t leave the day centre” rules. This is because many participants are unlikely to leave the day centre or walk inside staff offices as a matter of course, regardless of whether they are taking part in the MET-ID or not. Moreover, reducing the rules to three is consistent with the above findings that after being asked to learn the MET-ID rules for one minute, the mean number of rules spontaneously recalled was 2.02, which rose to 3.52 after the rules were cued. Therefore the use of just three rules might be a more manageable amount for the participants to hold in working memory. On top of this, introducing a post MET-ID measure asking how many of the rules the participants could remember might be useful to gain an understanding of the extent to which the participants were aware of the rules when they were undertaking the task to evaluate the validity of this scale further.

Nevertheless, to ensure that the MET-ID is still able to capture an appropriate range of performance, it might be beneficial to increase the number of tasks to offset the reduction in rules. As task completions appeared to be the most clinically useful scale (where some participants were able to complete all the tasks, some neglected some whereas others were unable to initiate any tasks at all) then this might provide further clinical and theoretical insights (see below) about the participant’s executive function. Moreover, increasing the tasks should not technically overload the participant’s working memory as more tasks would not necessarily mean having to switch between tasks and rules, thus avoiding a dual task or set shifting effect.

A further difficulty with the MET-ID was its reliance on receptive vocabulary, reading ability and general verbal abilities. Whilst, as stated, it would be very difficult to design a “pure” measure of executive function that doesn’t recruit other cognitive skills, the results showed that the MET-ID did correlate meaningfully with these variables. This is perhaps unavoidable. However, these findings may be pronounced because there were relatively few measures of executive function use to correlate with MET-ID performance. Unfortunately, due to the constraints involved in collecting the required quantity of data in the time available, additional measures of executive function could not pragmatically be employed in this study. However, such additional measures would be useful because two of the measures of executive function used here (the SPT and the DEX-IR) have not been widely explored in or designed for ID populations. This might explain why there was a lack of correspondence between these measures and the MET-ID. In contrast, the version of the TOLT used here was designed for an ID population and did correlate with the MET-ID. Therefore, other specific measures of executive function that have been developed for an ID population, such as the subtests from the CEFA, may be more appropriately compared with the MET-ID. Indeed, there are six explicit executive function measures on the CEFA which Ball et al. (2008) regard as tapping the functions of initiation, set shifting, working memory, efficient organisation of retrieval and recall, response inhibition, abstracting common principles and planning. Correlating the MET-ID with these measures may allow further exploration of the specific executive component(s) of the MET-ID, above and beyond receptive vocabulary, reading ability and general verbal abilities.

However, in doing this it is important to bear in mind that the subtests from the CEFA were not adapted with ecological validity in mind. Therefore, it would be useful for future research to adapt further laboratory based measures that can prove to be ecologically sensitive for ID populations. Adaptation of the subtests from the BADS-C may be a good starting point for this and work on the Hotel Task by Manly, Hawkins, Evans, Woldt, and Robertson (2002) in brain injured participants provides a good example of how this can be done. Such measures might provide a better way of assessing real world executive skills in the laboratory and be better suited for correlating with MET-ID performance.

Equally, for the MET-ID to truly back up its claims to be an ecologically valid assessment it needs to show correlations with measures of real world function. This study has found that the DEX-IR questionnaire may not be best suited for this purpose in an ID sample. Otherwise, the use of an alternate informant based questionnaire such as the Vineland Adaptive Behaviour Scales – Second Edition (Sparrow, Cicchettti, & Balla, 2005) may been a useful option as this assesses real world behaviours (e.g., communication, daily living, motor skills and socialisation) and is specifically designed for an ID population.

Additional problems with the study are evident in the sample used. As stated above, the participants were taken from what might have been the higher range of a potentially more globally impaired sample of people attending day services. This means that it is difficult to generalise the findings to those people with ID who do not attend day services who (by virtue of the fact that they don't attend day services) might be less cognitively impaired. On top of this, it is likely that it would be impractical for psychologists to assess people on the version of the MET-ID developed here (that is explicitly designed for day centres) for people who do not attend a day centre due to potential difficulties accessing the day centre for non members. Therefore, future research might focus on developing a version of the MET-ID than can be carried out within the examinees home as it is likely that many potential participants (whether attending a day service or not) would live in a residential setting in which they could be assessed, thus affording a potentially less biased sample and a plausibly more pragmatic environment.

### 4.6 Theoretical implications

The finding from this study that had the most overt theoretical implication is the robust relationship between MET-ID task completions and the TOLT. The reasons why the completion of specific tasks on the MET-ID would correlate with better performance on the TOLT therefore requires examination.

Shallice and Burgess (1991) suggested that performance on the MET requires use of the SAS to provide (a) goal articulation, (b) provisional plan formulation, (c) the creation of a mental ‘marker’ to schedule a delayed intention, (d) triggering of the marker (by an appropriate mental or physical event) and (e) a continual evaluation process that can lead to the development of subgoals to be achieved if the present circumstances prove to be unsatisfactory. In considering the more recent and detailed specifications of the SAS (Shallice, 2002; Shallice & Burgess 1996), its component processes have been fractionated into eight functions including spontaneous schema generation (implicitly knowing what to do), goal setting, adoption of processing mode (problem solving), episodic memory retrieval (drawing upon memory from past experiences), delayed intention marker realisation (remembering to do something in the future), working memory, monitoring and rejection of schema (Burgess & Alderman, 2004). Drawing upon this model it is possible to see how ability to initiate tasks on the MET could be attributed to the spontaneous schema generation (being able to implicitly develop a plan to achieve the task), goal setting (being able to adequately set oneself the goal of achieving a set task), adoption of processing mode (being able to problem solve any difficulties that may arise) and delayed intention marker realisation (being able to remember to go back to tasks that cannot be completed straight away) functions of the SAS. Accordingly, it is useful to understand how abilities in these areas would correlate with performance on the TOLT.

Aside from being a measure of “planning” (Shallice, 1982) the TOLT has also been conceptualised as a measure of goal-conflict resolution (Morris, Miotto, Feigenbaum, Bullock, & Polkey, 1997). For example, on the TOLT the participant has to consider a number of moves that will bring their disc arrangement closer to the required end state. Because it would place a high cognitive demand on working memory to calculate all possible sequences one must go through to get from the starting arrangement to the goal state, the participant must engage in a number of problem solving processes. Simon (1975) describes one of the most efficient strategies as being to divide the task into smaller subgoals and progressively move through the subgoals accordingly. Goal-subgoal conflict occurs when one has to make a move that, whilst needed to bring the discs to the required end state, necessitates that the participant makes a move that appears to take them away from the end goal state. Several examples of such moves are present on the TOLT used in this study (see Masson et al., 2010). Accordingly, successful performance on the TOLT could be explained via abilities in spontaneous schema generation (being able to implicitly develop a plan to make a correct move in the face of goal-subgoal conflict), goal setting (being able to develop effective sub-goals in order to bring them closer to the desired end state in the face of goal-subgoal conflict) and adoption of processing mode (being able to problem solve any difficulties that may arise the face of goal-subgoal conflict).

Thus, with the exception of delayed intention marker realisation, there is a correspondence in the spontaneous schema generation, goal setting and adoption of processing mode functions of the SAS that could plausibly explain the relationship between these two measures. It is these components of the SAS that could potentially be impaired in those unable to complete tasks on the MET-ID and performing poorly on the TOLT.

Alternately, a more parsimonious explanation could be found in Duncan's theory of goal neglect (Duncan, 1986, 1995; Duncan et al., 1995, 1996, 1997, 2008). As outlined above, this theory is based upon the importance of goal setting in human behaviour and how, in order to achieve set goals they must be divided into specific subgoals. Duncan suggests that people with executive difficulties tend to neglect such goals despite being aware of them. As both the MET-ID and the TOLT have the potential for goal neglect to be exposed (e.g., on the MET-ID via failing/neglecting tasks and on the TOLT via repeatedly breaking rules despite immediate verbal feedback from the examiner after each rule break), it is therefore plausible that failing to complete task on the MET-ID and poor performance on the TOLT could alternately be attributed to a neglect of specific goals. This interpretation is tentative however based on the fact that Duncan et al. (1996) define goal neglect as "the disregard of a task requirement even though it has been understood and remembered" (p. 257). Thus, firm conclusions based on this theory are limited because it is unclear as to the extent to which, at least on the MET-ID, task requirements were understood and remembered (based on the correlations between the MET-ID scales and receptive vocabulary, reading ability and VIQ, and the finding that participants found it difficult to remember and recall the rules).

## 4.7 Clinical implications

The clinical implications for this study can be divided into those related to assessment and intervention.

### 4.7.1 Assessment.

Newer neuropsychological assessments that can demonstrate ecological validity are needed and this is no exception for the development of executive function measures for people with ID (Burgess et al., 2006; Masson et al., 2010). Indeed, most current standard measures of executive function are typically too complex or rely too heavily on verbal skills for people with ID, resulting in floor effects (Masson et al., 2010). Accordingly, this study has shown that the use of auditory equipment can help circumvent some of the difficulties faced in accessing measures of EF for people with ID, where typically the verbal demands of the measures are too complex. As it stands, the MET-ID still needs further development, however this study has demonstrated that EF can be assessed outside of traditional office settings in people with ID and thus afford a richer formulation of how an individual’s EF may play out in their everyday life. These considerations should be taken into account for the future development of measures of executive function in people with ID.

#### 4.7.2 Intervention.

Ultimately, this study is about the assessment of executive function and therefore it is recognised that generalisations about the way in which such assessment may guide interventions may be vulnerable to inferences that are not substantiated by the data reported here. Nevertheless, the theoretical implications of the findings lead onto a number of potential clinical implications that are worthy of note. Based on the SAS model, Shallice and Burgess (1996) suggest that, if spontaneous schema generation fails to occur, the adoption of processing mode function (e.g., problem solving) can be used to devise an appropriate plan. They describe this as a process of problem formulation and orientation (where goal setting occurs), the deepening of a solving attempt, assessing if a solution is effective and finally a return to the first phase for checking and recapitulation. Thus, deficits in schema generation, goal setting and problem solving would lend themselves to interventions that explicitly address problem solving deficits. Accordingly, Problem Solving Therapy (PST; D'Zurilla & Nezu, 2007) attempts to address four major problem solving skills including (a) problem definition and formulation, (b) generation of alternative solutions, (c) decision making and (d) solution implementation and verification. Interventions based on this model have been shown to be effective at improving executive function in participants with brain injuries (Rath, Simon, Langenbahn, Sherr, & Diller, 2003; von Cramon, Matthes-von Cramon, & Mai, 1991). Accordingly, based on evidence that the application of problem solving therapy to difficulties faced in the social domain have been effective in people with ID (Lindsay, et al., 2011; Loumidis & Hill, 1997), it is conceivable that such a therapy may be successfully adapted for treating executive deficits in people with ID.

Alternately, as outlined above, Duncan’s theory of goal neglect has stimulated the development of GMT (Levine et al., 2000), which shares strong parallels with PST. Here GMT uses similar techniques to aid the participant in overcoming executive problems by getting them to (a) stop, (b) define what is to be done, (c) list the steps needed to do it, (d) learn the steps and (e) check it. This approach has also demonstrated promising results in people with brain injuries (Levine et al., 2000) and therefore may also offer useful implications for therapies designed to remediate executive impairments in people with ID.

Equally, there is also growing evidence that the MET can be a useful rehabilitation tool for improving executive difficulties in its own right, or at least in teaching people the use of coping strategies to foster improved MET performance (Alderman & Baker, 2009; Rand et al., 2009). Further research may substantiate the degree to which such training would generalise onto other real world executive tasks to investigate whether training on the MET-ID may be an additional potential means of improving cognitive skills in people with ID. Alternately, the MET-ID may serve as a useful means of evaluating outcome for any PST or GMT intervention.

## 4.8 Overall conclusion

The research hypotheses proposed in this study were not supported by the data as the MET-ID failed to correlate well with measures that purport to assess executive function deficits in real life. Moreover, MET-ID performance was strongly related to receptive vocabulary, reading comprehension and general verbal abilities. This limits the extent to which this preliminary version of the MET-ID can be generalised as an ecologically valid and specific assessment of executive function in people with ID. Thus, the MET-ID will need further development to make it a more accessible and psychometrically sound measure for an ID population. However, the MET-ID did share a robust relationship with the TOLT. This suggests that the two tasks may have a specific executive process (or set of executive processes) in common which can be attributed to the spontaneous schema generation, goal setting and adoption of processing mode (problem solving) functions of the SAS. Whist further research is needed, this suggests that specific interventions such as PST or GMT may prove fruitful at helping people with ID manage executive tasks in everyday life.

**References**

Adams, D., & Oliver, C. (2010). The relationship between acquired impairments of executive function and behaviour change in adults with Down syndrome. *Journal of Intellectual Disability Research, 53,* 393-405.

Alberoni, M., Baddeley, A., Della Sala, S., Logie, R., & Spinnler, H. (1992). Keeping track of a conversation: Impairments in Alzheimer's disease. *International Journal of Geriatric Psychiatry, 7,* 639-646.

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

Alderman, N., & Baker, D. (2009). Beyond the shopping centre: using the Multiple Errands Test in the assessment and rehabilitation of multi-tasking disorders. In M. Oddy, & A, Worthington. (Eds.), *Rehabilitation of executive disorders: A guide to theory and practice* (pp. 97-117). Oxford, UK: Oxford University Press.

Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands test. *Journal of the International Neuropsychological Society, 9,* 31-44.

Allport, A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In D. E. Meyer, & S. Kornblum (Eds.), *Attention and performance* (Vol. 14, pp. 183-218). Cambridge, MA: Bradford.

American Psychiatric Association (APA). (2000). *Diagnostic and Statistical Manual of the Mental Disorders (Fourth Edition-Text Revision, DSM-IV-TR).* Washington, DC: APA.

Andrés, P., & Van der Linden, M. (2000). Age related differences in supervisory attentional system functions. *Journal of gerontology: Psychological Sciences, 55B,* 373-380.

Andrés, P., & Van der Linden, M. (2001). Supervisory attentional system in patients with focal frontal lesions. *Journal of Clinical and Experimental Neuropsychology, 23,* 225-239.

Annett, M., Hudson, P. T. W., & Turner, A. (1974). The reliability of differences between the hands in motor skill. *Neuropsychologia, 12*, 527-531.

Armitage, S. G. (1946). An analysis of certain psychological tests used for the evaluation of brain injury. *Psychology Monographs*, *60, (whole no. 277)*.

Baddeley, A. D. (1966). The capacity for generating information by randomization. *Quarterly Journal of Experimental Psychology, 18,* 119-130.

Baddeley, A. D. (1986). *Working Memory*. Oxford University Press: Oxford.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A, 49,* 5-28.

Baddeley, A. D. (2001). Is working memory still working? *American Psychologist, 56,* 851-864.

Baddeley, A. D., Bressi, S., Della Sala, S., Logie, R., & Spinnler H. (1991). The decline of working memory in Alzheimer's disease: A longitudinal study. *Brain, 114,* 2521-2542.

Baddeley, A. D., Della Sala, S., Gray, C., Papagno, C., & Spinnler, H. (1997). Testing the central executive functioning with a pencil and paper test. In P. Rabbit (Eds.), *Methodology of frontal and executive function.* (pp. 61-80). East Sussex, UK: Psychology Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Eds.), *The Psychology of Learning and Motivation,* Vol. 8 (pp. 47-89). New York, US: Academic Press.

Baddeley, A. D., Logie, R., Bressi, S., Della Sala, S., & Spinnler, H. (1986). Dementia and working memory. *Quarterly Journal of Experimental Psychology*, *38A*, 603-618.

Baddeley, A. D., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition*, *7*, 212-230.

Ball, S. L., Holland, A. J., Treppner, P., Watson, P. C., & Huppert, F. A. (2008). Executive dysfunction and its association with personality and behaviour changes in the development of Alzheimer's disease in adults with Down syndrome and mild to moderate learning disabilities. *British Journal of Clinical Psychology, 47*, 1-29.

Baron, I. S. (2006). Behavioural assessment of the dysexecutive syndrome for children (BADS-C) by Emslie, H., Wilson, F. C., Burden, V., Nimmo-Smith, I., & Wilson, B. A. (2003). *Child Neuropsychology, 13*, 539-542.

Baum, C. M., Connor, L. T., Morrison, T., Hahn, M., Dromerick, A. W., & Edwards, D. F. (2008). Reliability, validity, and clinical utility of the Executive Function Performance Test : A measure of executive function in a sample of people with stroke. *American Journal of Occupational Therapy, 62*, 446-455.

Bechara, A., & Damiaso, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behaviour, 52,* 336-372.

Bennet, P. C., Ong, B., & Ponsford, J. (2005). Measuring executive dysfunction in an acute rehabilitation setting: Using the dysexecutive questionnaire (DEX). *Journal of the International Neurological Society, 11*, 376-385.

Benton, A. (1968). Differential behavioural effects in frontal lobe disease. *Neuropsychologia, 6*, 53-60.

Benton, A., Hamsher, K., & Sivan, A. (1978). *Multilingual aphasia examination: Manual of instructions – Third Edition*. Iowa City, IA: AJA Associates.

Bigler, E. D. (2001). Neuropsychological testing defines neurobehavioural significance of neuro-imaging defined abnormalities. *Achieves of Clinical Neuropsychology, 16,* 227-236.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioural and Brain Sciences, 29,* 109-125.

Borkowski, J. G., Benton, A L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia*, *5*, 135–140.

Bouquet, C. A., Bonnaud, V., & Gill, R. (2003). Investigation of supervisory attentional system functions in patients with Parkinson's disease using the Hayling task. *Journal of Clinical and Experimental Neuropsychology, 25,* 751-760.

Bourke, P. A., Duncan, J., & Nimmo-Smith, I. (1996). A general factor involved in dual-task performance decrement. *Quarterly Journal of Experimental Psychology, 49A,* 525-545.

Bressi, S., Miele, L., Bressi, C., & Astori, S. (1996). Deficit of central executive component of working memory in schizophrenia. *New Trends in Experimental and Clinical Psychiatry 12*, 243-252.

British Psychological Society. (2004). *Guidelines for minimum standards of ethical approval in psychological research.* Retrieved from http://www.bps.org.uk/document-download-area/document-download$.cfm?file_uuid=E6B70D7B-EE9D-EE0C-5F7A-F6CEB56874FB&ext=pdf

Brown, J. I., Nelson, M. J., & Denny, E. C. (1973). *Examiner's manual. The Nelson-Denny reading test.* Boston: Houghton-Miffin.

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

Burgess, P. W. (2010). Assessment of executive function. In J. Gurd, U. Kischka, & J. Marshall (Eds.), *The handbook of clinical neuropsychology* (pp. 349-369). Oxford, UK: Oxford University Press.

Burgess, P. W., & Alderman, N. (2004). Executive dysfunction. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical Neuropsychology: A practical guide to assessment and management for clinicians* (pp. 185-209). West Sussex, UK: Wiley.

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

Burgess, P.W., Alderman, N., Evans, J. J., Wilson, B. A., Emslie, H. & Shallice, T. (1996). Modified Six Element Test. In B. A. Wilson, N. Alderman, P. W. Burgess, H. Emslie, & J. J. Evans (Eds.), *Behavioural assessment of the dysexecutive syndrome*. Bury St. Edmunds, U.K.: Thames Valley Test Company.

Burgess, P. W., Alderman, N., Wilson, B. A., Evans, J. J., & Emslie, H. (1996). The Dysexecutive Questionnaire. In B. A. Wilson, N. Alderman, P. W. Burgess, H. Emslie, & J. J. Evans (Eds.), *Behavioural assessment of the dysexecutive syndrome*. Bury St Edmunds, U.K.: Thames Valley Test Company.

Burgess, P. W., Alderman, N., Volle, E., Benoit, R. G., & Gilbert, S. J. (2009). Mesulam's frontal lobe mystery re-examined. *Restorative Neurology and Neuroscience, 27,* 493-506.

Burgess, P. W., Alderman, N., Forbes, C., Lacy Costello, A., Coates, L, M-A., Dawson, D, D ... Channon, S. (2006). The case for the development of and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society, 12,* 194-209.

Burgess, P. W., & Robertson, I. H. (2002). Principles of the rehabilitation of frontal lobe function. In D. T., Stuss & R. T. Knight (Eds.), Principles of frontal lobe function (pp. 557-572). Oxford: Oxford University Press.

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

Burgess, P. W., & Shallice, T. (1996b). Bizarre responses, rule detection and frontal lobe lesions. *Cortex, 32,* 241-259.

Burgess, P.W., & McNeil, J. E. (1999). Content-specific confabulation. *Cortex, 35,* 163-182.

Bush, H, M. (2011). *Biostatistics: An applied introduction for the public health practitioner.* New York: Delmar Cengage Learning.

Carr, A., & O'Reilly, G. (2007). Diagnosis, classification and epidemiology. In A. Carr, G. O'Reilly, P. Noonan Walsch, & C, McEvoy (Eds.), *The handbook of intellectual disability and clinical psychology practice* (pp. 3-49). East Sussex: Routledge

Cavedini, P., Riboldi, G., Keller, R., D'Annucci, A., & Bellodi, L. (2002). Frontal lobe dysfunction in pathological gambling patients. *Biological Psychiatry, 51,* 334-341.

Chan, R. C. K. (2001). Dysexecutive symptoms among a non-clinical sample: A study with the use of the Dysexecutive Questionnaire. *British Journal of Psychology, 92,* 551-565.

Chan, R. C. K., Shum, D., Toulopolou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology, 23,* 201-216.

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

Chevignard, M. P., Pillon, B., Pradat-Diehl, P., Taillefer, C., Rousseau, S., Le Bras, C., & Dubois B. (2000). An ecological approach to planning dysfunction: Script execution. *Cortex, 36,* 649-669.

Chevignard, M. P., Taillefer, C., Picq, C., Poncet, F., Noulhiane, M. & Pradat-Diehl, P. (2008). Ecological assessment of the dysexecutive syndrome using execution of a cooking task. *Neuropsychological Rehabilitation, 18,* 461-485.

Clare, L., & Woods, R. T. (2001). *Cognitive rehabilitation in dementia: Special issue of the journal of neuropsychological rehabilitation.* East Sussex: Psychology Press.

Dalrymple-Alford, J. C., Kalders, A. S., Jones, R. D., & Watson, R. W. (1994). A central executive deficit in patients with Parkinson’s disease. *Journal of Neurology, Neurosurgery and Psychiatry, 57,* 360 – 367.

Danielson, H., Henry, L., Rönnberg, J., & Nilsson, L. G. (2010). Executive functions in individuals with intellectual disability. *Research in Developmental Disabilities, 31,* 1299-1304.

Data Protection Act. (1998). *Chapter 29.* London: HMSO.

Dawson, D. R., Anderson, N. D., Burgess, P. W., Cooper, E., Krpan, K. M., & Stuss, D. T. (2009). Further development of the Multiple Errands Test: Standardised Scoring, Reliability, and Ecological Validity for the Baycrest Version. *Archives of Physical Medicine Rehabilitation, 90,* S41-S51.

Demakis, G. J. (2003). A meta-analytic review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. *Neuropsychology, 17,* 255-64.

Demakis, G. J. (2004). Frontal lobe damage and tests of executive processing: A meta-analysis of the Category Test, Stroop Test, and Trail-Making Test. *Journal of Clinical and Experimental Neuropsychology, 26,* 441-450.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001) *Delis-Kaplan Executive Function System*. San Antonio, TX: The Psychological Corporation.

Della Sala, S., Foley, J. A., Beschin, N., Allerhand, M., & Logie, R. H. (2010). Dual task performance using a pencil and paper test: Normative data. *Archives of Clinical Neuropsychology, 25,* 410-419.

Duncan, J. (1986). Disorganisation of behaviour after frontal lobe damage. *Cognitive Neuropsychology*, *2*, 271–290.

Duncan, J. (1995). Attention, intelligence and the frontal lobes. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.

Duncan, J., Burgess, P. W., & Emslie, H. (1995). Fluid Intelligence after frontal lobe lesions. *Neuropsychologia, 33,* 261-268.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: the organisation of goal directed behaviour. *Cognitive Psychology, 30,* 257–303.

Duncan, J., Johnson, R., Swales, M., & Freer, C. (1997). Frontal lobe deficits after head injury: Unity and diversity of function. *Cognitive Neuropsychology, 14*, 713–741.

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., ... Nimmo-Smith, I. (2008). Neglect and Spearman's *g*: Competing Parts of a complex task. *Journal of Experimental Psychology: General, 137,* 131-148.

Dunn, L. M., Dunn, L. M., Styles, B., & Sewell, J. (2009). *British Picture Vocabulary Scale – Third Edition (BPVS-III).* Windsor, Berks: NFER Nelson.

Dymond, S., Bailey, R., Willner, P., & Parry, R. (2010). Symbol labelling improves advantageous decision-making on Iowa Gambling Task in people with intellectual disabilities. *Research in Developmental Disabilities, 31,* 536-544.

D'Zurilla, T. J., & Nezu, A. M. (2007). Problem solving therapy: A positive approach to clinical intervention. New York: Springer.

Emslie, H., Wilson, F. C., Burden, V., Nimmo-Smith, I., & Wilson, B. A. (2003).*Behavioural assessment of the dysexecutive syndrome for children* (BADS-C). Bury St. Edmunds, UK: Thames Valley Test Company.

Eslinger, P. J., & Damasio, A. R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology, 35,* 1731-1741.

Field, A. (2009). *Discovering statistics using SPSS – Third edition.* London: Sage.

Fisher, A. G., & Bray Jones, K. (2010). *Assessment of Motor and Process Skills. Vol. 1: Development, Standardization, and Administration Manual – Seventh Edition.* Fort Collins, CO: Three Star Press.

Foley, J. A., Cantagallo, A., Della-Sala, S., & Logie, R. H. (2010). Dual task performance and post traumatic brain injury. *Brain Injury, 24,* 851-858.

Glenn, S., & Cunningham, C. (2005). Performance of young people with downs syndrome on the Leiter–R and the British picture vocabulary scales*. Journal of Intellectual Disability Research, 49*, 239-244.

Goldstein, L. H., Bernard, S., Fenwick, P. B. C., Burgess, P. W., & McNeil, J. (1993). Unilateral frontal lobectomy can produce strategy application disorder. *Journal of Neurology, Neurosurgery and Psychiatry, 56,* 274-276.

Goodman, R. (1999). The Extended Version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40,* 791-799.

Green, J. D. W., Hodges, J. R., Baddeley, A. D. (1995). Autobiographical memory and executive function in early dementia of Alzheimer type. *Neuropsychologia, 33,* 1647-1670.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Odessa, FL: Psychological Assessment Resources.

Hegarty, M., Shah, P., & Miyake, A. (2000). Constraints on using the dual task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition, 28,* 376-385.

Hinkin, C. H., Castellon, S. A., & Hardy, D. J. (2000). Dual task performance in HIV-1 infection. *Journal of Clinical and Experimental Neuropsychology, 22,* 16-24.

Huntley, J. D., & Howard, R. J. (2009). Working memory in early Alzheimer's disease: a neuropsychological review. *International Journal of Geriatric Psychiatry, 25,* 121-132.

Institute for Personality and Ability Testing. (1973). *Measuring intelligence with the Culture Fair tests.* Champaign, IL: The Institute for Personality and Ability Testing.

Jarvis, P. E., & Barth, J. T. (1984). *Halstead-Reitan Test Battery: An interpretive guide*. Odessa, FL: Psychological Assessment Resources.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review*, *17,* 213-233.

Junghoon, K., Glahn, D. C., Neuechterlein, K. H., & Cannon, T. D. (2004). Maintenance and manipulation of information in schizophrenia: Further evidence for impairment in the central executive component of working memory. *Schizophrenia Research, 68,* 173-187.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual differences perspective. *Psychonomic Bulletin and Review, 9,* 637-671.

Kaufman, A. S., & Lichtenberger, E, O. (1999). *Essentials of WAIS-III assessment*. New York: Wiley.

Kittler, P. M., Krinsky-McHale, S. J., Devenny, D. A., & Conners, F. (2008). Dual-Task processing as a measure of executive function: A comparison between adults with Williams and Down Syndromes. *American Journal on Mental Retardation, 113,* 117-132.

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.

Kofler, M. J., Rapport, M. D., Bolden, J., Sarver, D. E., & Raiker, J. S. (2009). ADHD and working memory: The impact of central executive deficits and exceeding storage/rehearsal capacity on observed inattentive behaviour. *Journal of Abnormal Child Psychology, 38,* 149-161.

Lamberts, F. L., Evans, J. J., & Spikman, J. M. (2009). A real-life, ecologically valid test of executive functioning: The executive secretarial task. *Journal of Clinical and Experimental Neuropsychology, 32,* 56-65.

Lanfranchi, S., Baddeley, A., Gathercole, S., & Vianello, R. (2011). Working memory in Down Syndrome: Is there a dual task deficit. *Journal of Intellectual Disability Research,* Advance online publication. doi: 10.1111/j.1365-2788.2011.01444.x

Lange, K. W., Shakian, B. J., Quinn, N. P., Marsden, C. D., & Robbins, T. W. (1995). Comparison of executive and visuospatial memory function in Huntingdon's disease and dementia of Alzheimer's type matched for degree of dementia. *Journal of Neurology, Neurosurgery and Psychiatry, 58,* 598-606.

Levine, B., Robertson, I. H., Clare, L., Carter, G., Hong, J., Wilson, B. A., ... 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.

Lindsay, W. R., Hamilton, C., Moulton, S., Scott, S., Doyle, M., & McMurran, M. (2011). Assessment and treatment of social problem solving in offenders with intellectual disability. *Psychology, Crime & Law, 17,* 181-197.

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test (Form 6).* Itasca, IL: Riverside.

Loumidis, K., & Hill, A. (1997). Social problem solving groups for adults with learning disabilities. In B. Stenfert Kroese, D. Dagnan, & K, Loumidis (Eds.), *Cognitive-Behaviour therapy for people with learning disabilities* (pp. 86-109). East Sussex: Routledge.

Luria, A. R. (1966). *Higher cortical functions in man*. London: Tavistock.

Malapani, C., Pillon, B., Dubois, B., & Agid, Y. (1994). Impaired simultaneous cognitive task performance in Parkinson's disease: A dopamine-related dysfunction. *Neurology, 44*, 319–326.

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

Marczewski, P., Van der Linden, M., & Larøi, F. (2001). Further investigation of the supervisory attentional system in schizophrenia: Planning inhibition and rule abstraction. *Cognitive Neuropsychiatry, 6,* 175-192.

Masson, J. D., Dagnan, D., & Evans, J. (2010). Adaptation and validation of the Tower of London test of planning and problem solving in people with intellectual disabilities. *Journal of Intellectual Disabilities Research, 54,* 457-467.

McDowell, S., Whyte, J., & D'Esposito, M. (1997). Working memory impairments in traumatic brain injury: evidence from a dual task paradigm. *Neuropsychologia, 35,* 1341-1353.

McGeorge, P., Philips, L. H., Crawford, J. R., Garden, S. E., Della Sala, S., Milne, A. B., ... & Carpenter, J. S. (2001). Using Virtual environments in the assessment of executive dysfunction. *Presence, 4,* 375-383.

McKenna, P., & Warrington, E. K. (1983). *Graded naming test*. Windsor, UK: NFER-Nelson.

McPheeters, H., L. (1984). Statewide mental health outcome evaluation: A perspective of two southern states. *Community Mental Health Journal*, *20*, 44–55.

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 1,* 172-175.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Morris, R. G., Miotto, E. C., Feigenbaum, P. B., Bullock, P., & Polkey, C. E. (1997). The effect of goal-subgoal conflict on planning ability after frontal- and temporal-lobe lesions in humans. *Neuropsychologia, 35,* 1147-1157.

Must, A., Szabo, Z., Bodi, N., Szasz, A., Janka, Z., & Keri, S. (2006). Sensitivity to reward and punishment and the prefrontal cortex in major depression. *Journal of Affective Disorders, 90,* 209–215.

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

Nelson, H. E. (1982). *National adult reading test manual*. Windsor, U.K.: NFER-Nelson.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (pp. 1-18). New York, USA: Plenum.

Norris, G., & Tate, R. L. (2000). The Behavioural assessment of the dysexecutive syndrome (BADS): Ecological, Concurrent and Construct Validity. *Neuropsychological Rehabilitation, 10,* 33-45.

Oddy, M., & Worthington, A. (2009). *Rehabilitation of executive disorders: A guide to theory and practice*. Oxford, UK: Oxford University Press.

Oscar-Berman, M., & Marinković, K. (2007). Alcohol: Effects on neurobehavioral functions and the brain. *Neuropsychology Review, 17,* 239-257.

Parkin, A. J. (1998). The central executive does not exist. *Journal of the International Neuropsychological Society, 4,* 518-522.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220-244.

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

Pennington, E. A. (2006). *Development of a simplified version of the Multiple Errands Test for use on a hospital ward* (Unpublished doctoral thesis). University of Leicester, Leicester.

Perry, R. J., & Hodges, J. R. (1999). Attention and executive deficits in Alzheimer's disease. *Brain, 122,* 383-404.

Petrides, M., & Milner, B. (1982). Deficits on subject ordered tasks after frontal and temporal lobe lesions in man. *Neuropsychologia, 20*, 249–264.

Porteus, S.D. (1965). *Porteus Maze Test. Fifty years' application*. New York: Psychological Corporation.

Rainville, C., Amieva, H., Lafont, S., Dartigues, J. F., Orgigizo, J. M., & Fabrigoule, C. (2002). Executive function deficits in patients with dementia of the Alzheimer's type: A study with a Tower of London task. *Achieves of Clinical Neuropsychology, 17,* 513-530.

Rand, D., Basha-Abu Rukan, S., Weiss, P. L., & Katz, N. (2009). Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychological Rehabilitation, 19,* 583-602.

Rand, D., Weiss, P. L., & Katz, N. (2009). Training multitasking in a virtual supermarket: A novel intervention after stroke. *American Journal of Occupational Therapy, 63,* 535-542.

Rath, J. F., Simon, D., Langenbahn, D. M., Sherr, L. R., & Diller, L. (2003). Group treatment of problem-solving deficits in outpatients with traumatic brain injury: A randomised outcome study. *Neuropsychological Rehabilitation, 13,* 461-481.

Raven, J. C., Court, J. H., & Raven, J. (1988). *Manual for Ravens Progressive Matrices and Vocabulary Scales.* London: H. K. Lewis.

Ricci, C., & Blundo, C. (1990). Perception of ambiguous figures after focal brain lesions. *Neuropsychologia, 28*, 1163–1173.

Robbins, T. W. (1998). Dissociating executive functions of the prefrontal cortex. In A. C. Roberts, T. W. Robbins, & L. Weiskrantz (Eds.), *The Prefrontal Cortex: Executive and Cognitive Functions* (pp. 117–130). Oxford: Oxford University Press.

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

Robertson, I. H. (1996). *Goal Management Training: A clinical manual*. Cambridge, U.K.: PsyConsult.

Rylander, G. (1939). Personality changes after operation on the frontal lobes. *Acta Psychiatrica Neurologica*, Supplement No. 30.

Salthouse, T. A., Fristone, N. M., Lineweaver, T. T., & Coon, V.C. (1995). Aging of attention: Does the ability to divide decline? *Memory and Cognition, 23,* 59-71.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B, 289,* 199-209.

Shallice, T. (2002). Fractionation of the supervisory system. In D. T. Stuss, & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 261-277). Oxford: Oxford University Press.

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

Shallice, T., & Burgess, P.W. (1996). The domain of supervisory processes and temporal organisation of behaviour. *Philosophical Transactions of the Royal Society of London B, 351,* 1405-1412.

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

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7,* 268–288.

Sparrow, S. S., Cicchettti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behaviour Scales – Second Edition (Vineland-II).* Minneapolis, Minnesota: Pearson Assessment.

Spearman, C. E. (1927). *The abilities of man*. New York: Macmillan.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms and commentary (Second Edition).* New York: Oxford University Press.

Spinnler, H., Della Sala, S., Bandera, R., & Baddeley, A. (1988). Dementia, ageing and the structure of human memory. *Cognitive Neuropsychology, 5,* 193-211.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Stuss, D. T., & Alexander, M. P. (2007). Is there a dysexecutive syndrome? *Philosophical Transactions of the Royal Society of London B, 362,* 901-951.

Stuss, D. T., & Benson, D. F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin, 95,* 3-28.

Teuber, H. L. (1972). Unity and diversity of frontal lobe functions. *Acta Neurobiologiae Experimentalis, 32,* 615-656.

Tranel, D., Hathaway-Nepple, J., & Anderson, S. W. (2007). Impaired behaviour on real-world tasks following damage to the ventromedial prefrontal cortex. *Journal of Clinical and Experimental Neuropsychology, 29,* 319-332.

Verghese, J., Kuslansky, G., Holtzer, R., Katz, M., Xiaonan, X., Buschke, H., & Pahor, M. (2007). Walking while talking: Effect of task prioritisation in the elderly. *Archives of Physical Medicine and Rehabilitation, 88,* 50-53.

Von Cramon, D., Matthes-von Cramon, G., & Mai, N. (1991). Problem-solving deficits in brain injured patients: A therapeutic approach. *Neuropsychological Rehabilitation, 1,* 45-64.

Warrington, E. K. (1984). *Recognition memory test*. Windsor, UK: NFER-Nelson.

Warrington, E. K., & Taylor, A. M. (1973). The contribution of the right parietal lobe to Object recognition. *Cortex, 9*, 152–164.

Warrington, E. K., James, M., & Maciejewski, C. (1986). The WAIS as a lateralising and localising diagnostic instrument: A study of 656 patients with unilateral cerebral lesions. *Neuropsychologia, 24,* 223-239.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1998) *Wechsler Adult Scale of Intelligence – Third Edition (WAIS-III)*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1999) *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2005). *Wechsler Individual Attainment Test – Second UK Edition (WIAT-II UK).* London, UK: Psychological Corporation.

Wechsler, D., & Stone, C. P. (1945). *Wechsler Memory Scale.* New York: Psychological Corporation.

Welford, A. T. (1952). The “psychological refractory period” and the timing of high speed performance – A review and a theory. *British Journal of Psychology, 43,* 2-19.

Wilkins, A., Shallice, T., & McCarthy, R. (1987). Frontal lesions and sustained attention. *Neuropsychologia, 25*, 359–365.

Willner, P., Bailey, R., Parry, R., & Dymond, S. (2010a). Evaluation of executive functioning in people with intellectual disabilities. *Journal of Intellectual Disabilities Research, 54,* 366-379.

Willner, P., Bailey, R., Parry, R., & Dymond, S. (2010b). Evaluation of the ability of people with intellectual disabilities to ‘weigh up’ information in two tests of financial reasoning. *Journal of Intellectual Disabilities Research, 54,* 380-391.

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

Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome, *Neuropsychological Rehabilitation, 8,* 213-228.

Wilson, R. S., Baker, L. D., Fox, J. H., & Kaszniak, A. (1983). Primary memory and secondary memory in dementia of Alzheimer type. *Journal of Clinical Neuropsychology, 5,* 337-344.

Wykes, T., & Reeder, C. (2005). *Cognitive remediation therapy for schizophrenia; Theory and practice.* East Sussex: Routledge

# Appendices

**Appendix A - Exercise sheet for the Multiple Errands Test - Hospital Version**

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 2T

**Appendix B - Information Sheet for the ID Sample**

School of Medicine, Health Policy and Practice

**University of East Anglia**
Norwich NR4 7TJ England

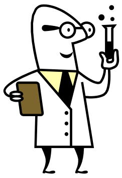
*Telephone*
01603 593310
*Fax*
01603 591132
*Email*
t.steverson@uea.ac.uk

**Information sheet ID sample (Version 2 – 11.11.2010)**

## <u>Development of a new test of multitasking</u>

### Who are the researchers?

- Tom Steverson
- Dr Peter Langdon
- Dr Anna Adlam

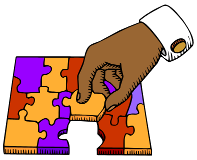


### What is it about?

- Some people have difficulties doing lots of different things at the same time.
- We are trying to find out how this might affect the daily lives of people with learning difficulties to find better ways of helping them in their daily lives.
- The NHS Norfolk Research and Development department and the Cambridgeshire 4 Research Ethics Committee (ref 10/H0305/75) have approved this study.

**Do I have to take part in this research?**

- No, it's up to you if you want to take part. If you don't want to take part this will not affect your care at the day centre or any treatment you receive from the NHS or social services.
- If you say YES, Tom will come and speak to you and answer any questions you might have.
- If you want to take part, he will ask you to do some puzzles and quizzes.
- He will also ask a member of staff to complete a questionnaire about you. However, he will make sure you are happy for him to do this and show you the questionnaire.



**How long will it all take?**

- About one hour on two different days.
- I will make sure that it does not interfere with your daily activities.

**Will my information be kept secret?**

- Yes, the information will be kept under lock and key and treated as confidential under the data protection act (1998).
- I will write about the information I have collected but I will make sure that there are no personal details about you.

**Are there any bad things that could happen?**

- No
- Some of the tasks might be a little difficult, but they are designed to be like that.
- If you are upset in any way, you can stop the tasks and talk to Tom or someone else.

**What if I am unhappy about the research?**

- You can stop at any point during the study and tell Tom that you no longer wish to take part.
- Any information collected about you will then be destroyed.

**Then…**

- You can talk to Tom or someone you know about it.
- You can make a complaint to the University of East Anglia or the NHS.
- You will be given information on how to complain.
- You may want someone you know to help you with this.
- If you are harmed, you may be able to take legal action - if this is caused by the study.
- If you harm yourself in a way that is not caused by taking part in the study then we won't be able to help you with any legal action.

**Contacts**

- If you want more information or wish to complain, you or your key worker can call UEA (01603 593310, Monday to Friday) and ask to speak to:

Tom Steverson or
Dr Peter Langdon (Clinical Psychologist) or
Dr Anna Adlam (Clinical Psychologist).

- Or write to any of these people at:

School of Medicine, Health Policy and Practice
University of East Anglia
Norwich
NR4 7TT

Any questions?

## Appendix C - Informed Consent Form for the ID Sample

School of Medicine, Health Policy and Practice

University of East Anglia

**University of East Anglia**
Norwich NR4 7TJ England

*Telephone*
01603 593310

*Fax*
01603 591132

*Email*
t.steverson@uea.ac.uk

**Informed consent from ID sample (Version 2 – 11.11.2010)**

Would you like a carer/key worker to be present?

Please tick the box if you agree with the sentence:

| | |
|---|---|
| I understand the information sheet for this study explained to me by ........................................................... | |
| I have asked any questions I have wanted to. | |
| I understand that I don't have to take part in this research and I can leave at any time without giving a reason. | |
| I understand that it will not affect my care at the day centre if I choose not to take part. | |
| I understand that all information collected about me is confidential. | |
| I am happy for the staff to complete a questionnaire about me and I have seen this questionnaire. | |
| I agree to take part in the research. | |
| I agree that the above questions have been answered adequately | |

Name of Participant________________

Signature____________________

Date________________________

Name of Researcher_________________

Signature______________________

Date__________________________

## Appendix D - Demographic Data Collection Sheet for the ID Sample

School of Medicine, Health Policy and Practice

University of East Anglia
**University of East Anglia**
Norwich NR4 7TJ England

*Telephone*
01603 593310
*Fax*
01603 591132
*Email*
t.steverson@uea.ac.uk

**Demographic data collection sheet ID sample (Version 1 – 06.10.2010)**

Participant identification number_______________________

Gender________________________________________________

1. What is your date of birth?
2. Have you ever had a head injury or been diagnosed with dementia or schizophrenia?
3. Are you currently taking part in other research?
4. How well do you know the day centre on a scale of 0-4 (0 = not very well, 2 = average, 4 = very well).

**Appendix E - Information Sheet for Staff Sample**

School of Medicine, Health Policy and Practice

University of East Anglia

**University of East Anglia**
Norwich NR4 7TJ England

*Telephone*
01603 593310
*Fax*
01603 591132
*Email*
t.steverson@uea.ac.uk

**Information sheet staff version (Version 2 – 11.11.2010)**

## Development of a new test to examine multitasking

### Who are the researchers?
- Tom Steverson
- Dr Peter Langdon
- Dr Anna Adlam

### What is it about?
- Some people with learning difficulties might have problems multitasking.
- We are trying to find out how this might affect the daily lives of people with learning difficulties by asking staff to fill out a 20 item questionnaire about a service user at the day centre.
- We hope that the work will help provide better ways of helping people with learning difficulties in their daily lives.
- The study has been reviewed and given favourable opinions by Cambridgeshire 4 Research Ethics Committee (ref number 10/H0305/75) and the NHS Norfolk Research and Development department.

### Do I have to take part in this research?
- No, it's up to you if you want to take part or not.
- If you agree, I will ask you to fill out a questionnaire about one of the service users at the day centre.
- The service user will have agreed to take part in the research and given consent to being rated on the questionnaire by a member of staff.

**How long will it all take?**

- The questionnaire has only 20 questions so the whole process should only take about 10 minutes.

**Is the information I give confidential?**

- Yes, all data is anonymised and the service user will not be able to see the answers you give.
- The information will be kept under lock and key and treated as confidential under that data protection act (1998).
- I will write up the information I have collected in my research but I will make sure that there are no personal details about you.

**What if I am unhappy about the research?**

- You can stop at any point during the study and withdraw your participation.
- Any information you gave will then be destroyed.

**If you remain unhappy about the research**

- You can make a complaint to the University of East Anglia or the NHS.
- You will be given information on how to complain.
- If you are harmed, you may be able to take legal action.

**Contacts**

- If you want more information or wish to complain, you can call UEA (01603 593310, Monday to Friday) and ask to speak to:

Tom Steverson or
Dr Peter Langdon (Clinical Psychologist) or
Dr Anna Adlam (Clinical Psychologist).

- Or write to either of these people at:

School of Medicine, Health Policy and Practice
University of East Anglia
Norwich
NR4 7TT

Any questions?

## Appendix F - Informed Consent Form for Staff Sample

School of Medicine, Health Policy and Practice

University of East Anglia

**University of East Anglia**
Norwich NR4 7TJ England

*Telephone*
01603 593310
*Fax*
01603 591132
*Email*
t.steverson@uea.ac.uk

**Informed consent form staff version (Version 2 – 11.11.2010)**

Please tick the box if you agree with the sentence:

| | |
|---|---|
| I understand the information sheet for this study explained to me by Tom Steverson. | |
| I have asked any questions I have wanted to. | |
| I understand that I don't have to take part in this research and I can leave at any time without giving a reason. | |
| I understand that all information I give is confidential. | |
| I agree to take part in the research. | |
| I agree that the above questions have been answered adequately. | |

Name of Participant_________________

Signature_____________________

Date_________________________

Name of Researcher________________

Signature_____________________

Date_________________________

## Appendix G - Ethical Approval

NHS

**National Research Ethics Service**

**Cambridgeshire 4 Research Ethics Committee**

Victoria House
Capital Park
Fulbourn
Cambridge
CB21 5XB

Telephone: 01223 597685
Facsimile: 01223 597645

23 November 2010

Mr Tom Steverson
Trainee Clinical Psychologist
School of Medicine, Health Policy and Practice
University of East Anglia
Norwich
NR4 7TJ

Dear Mr Steverson

| | |
|---|---|
| **Study Title:** | **Ecological assessment of the supervisory attentional system in people with intellectual disabilities.** |
| **REC reference number:** | **10/H0305/75** |

Thank you for your letter of 15 November 2010, responding to the Committee's request for further information on the above research, and submitting revised documentation.

The further information was considered, in correspondence by a sub-committee of the REC. A list of the sub-committee members is attached.

**Confirmation of ethical opinion**

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation, as revised, subject to the conditions specified below.

**Ethical review of research sites**

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

**Conditions of the favourable opinion**

The favourable opinion is subject to the following conditions being met prior to the start of the study.

<u>Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.</u>

For NHS research sites only, management permission for research ("R&D approval") should be obtained from the relevant care organisation(s) in accordance with NHS research governance arrangements. Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

*Where the only involvement of the NHS organisation is as a Participant Identification Centre (PIC), management permission for research is not required but the R&D office should be notified of the study and agree to the organisation's involvement. Guidance on procedures for PICs is available in IRAS. Further advice should be sought from the R&D office where necessary.*

*Sponsors are not required to notify the Committee of approvals from host organisations.*

**It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).**

**Approved documents**

The final list of documents reviewed and approved by the Committee is as follows:

| *Document* | *Version* | *Date* |
|---|---|---|
| Investigator CV | Tom Steverson | 14 August 2010 |
| Protocol | 1 | 08 October 2010 |
| MET-ID Exercise Sheet | 1 | 06 October 2010 |
| CV for Peter Langdon | | |
| REC application | 62013/155712/1/379 | 06 October 2010 |
| Covering Letter | Tom Steverson | 08 October 2010 |
| Questionnaire: Dex Questionnaire | | |
| GP/Consultant Information Sheets | 1 | 06 October 2010 |
| Participant Information Sheet: ID Sample | 2 | 11 November 2010 |
| Response to Request for Further Information | Tom Steverson | 15 November 2010 |
| Participant Information Sheet: Staff version | 2 | 11 November 2010 |
| Participant Consent Form: ID sample | 2 | 11 November 2010 |
| Participant Consent Form: Staff version | 2 | 11 November 2010 |
| Demographic data collection sheet ID sample | 1 | 06 September 2010 |
| CV for Anna Adlam | | |
| Demographic data staff version | 1 | 06 September 2010 |
| Evidence of insurance or indemnity | UEA | 02 September 2010 |

**Statement of compliance**

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

**After ethical review**

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk.

| 10/H0305/75 | **Please quote this number on all correspondence** |
|---|---|

With the Committee's best wishes for the success of this project

Yours sincerely

pp. NSStorey

**Dr Leslie Gelling**
**Chair**

Email: Nicky.Storey@eoe.nhs.uk

*Enclosures:* *List of names and professions of members who were present at the meeting and those who submitted written comments*

"After ethical review – guidance for researchers"

*Copy to:* Tracy Moulton
Research, Enterprise & Engagement Office
The Registry
University of East Anglia
Norwich
NR4 7TJ

## Appendix H - Research Governance Approval

NHS
Norfolk

Ref: 2010LD02

**Research & Development**
NHS Norfolk
Lakeside 400
Old Chapel Way
Broadland Business Park
Thorpe St Andrew
Norwich
NR7 0WG

Tel: 01603 257283
Fax: 01603 257292
E-mail: paul.mills@norfolk.nhs.uk
www.norfolk.nhs.uk/research

Tom Steverson
School of Medicine, Health Policy & Practice
University of East Anglia
Earlham Road
Norwich
NR4 7TJ

17 December 2010

Dear Tom,

**Re: 2010LD02. Ecological assessment of the supervisory attentional system in people with intellectual disabilities.**

Further to your submission of the above project to the R&D office at NHS Norfolk your project has now been reviewed and all the mandatory research governance checks have been satisfied. I am therefore pleased to inform you on behalf of **Norfolk Community Health & Care NHS Trust** that NHS permission (R&D approval) has now been granted for your study to take place at the following sites:

- Norfolk Community Health & Care NHS Trust

Please note that NHS Permission is granted on the basis of the information supplied in the research governance submission, if anything subsequently comes to light that would cast doubts upon, or alter in any material way, any information contained in the original application, or a later amendment application there may be implications for continued NHS Permission.

Please note the following conditions of approval:

- Please contact the Head of LD Provision of Community Services at Norfolk County Council prior to contacting day centre managers
- The R&D office at NHS Norfolk understands that <u>only</u> anonymous data will be held at your home address and any documentation which may be identifiable will be securely stored at the University of East Anglia

Please note it is your responsibility to ensure that these conditions are disseminated to all parties involved in this project.

You may now begin your study at the above sites.

Stonewall
DIVERSITY CHAMPION

POSITIVE ABOUT DISABLED PEOPLE

Chair: Sheila Childerhouse
Chief Executive: Andrew Morgan

Visit our website: www.norfolk.nhs.uk

I have enclosed two copies of the Standard Terms and Conditions of Approval. Please sign and return one copy to the R&D office at the above address. Failure to return the standard terms and conditions may result in NHS permission being revoked.

**Please note, under the agreed standard terms and conditions you must inform the R&D Office at NHS Norfolk of any proposed changes to this study, whether minor or substantial, and to keep the Committee updated on progress. Please note also, if you wish to extend approval to any sites other than those listed above you must apply for this through the relevant R&D office.**

If you have any queries regarding this or any other project please contact Paul Mills, R&D Officer, at the above address. Please note, the reference number for this study is **2010LD02** and this should be quoted on all correspondence.

The following documents were reviewed:

Letter of Favourable Opinion from Cambridgeshire 4 REC, dated 23rd November 2010
- Protocol, Version 1, 6th October 2010
- Investigator CV – Tom Steverson
- Investogator CV – Peter Langdon
- Investigator CV – Anna Adlam
- MET-ID Exercise Sheet, Version 1, 6th October 2010
- DEX Questionnaire
- GP Letter, Version 1, 6th October 2010
- Participant Information Sheet – ID Sample, Version 2, 11th November 2010
- Participant Consent Form – ID Sample, Version 2, 11th November 2010
- Participant Information Sheet – Staff, Version 2, 11th November 2010
- Participant Consent Form – Staff, Version 2, 11th November 2010
- Demographic Data Collection Sheet – ID Sample, Version 1, 6th October 2010 (see note)
- Evidence of Insurance/Indemnity, 2nd September 2010
- Response to Request for Further Information, 15th November 2010

Notification to Cambridgeshire 4 REC of Minor Amendment, dated 30th November 2010
- Letter to Cambridgeshire 4 REC, 30th November 2010
- Demographic Data Collection Sheet – Staff, Version 2, 30th November 2010
- Site Information Sheet, Version 1, 30th November 2010

Other Documents Reviewed
- Signed R&D Form, Lock Code 62013/165733/14/327
- Signed SSI Form, Lock Code 62013/148266/6/241/74008/189201

Note
- It is noted that the Demogrpahic Data Collection Sheet is marked on the REC approval letter as Version 1, 6th September 2010, whereas the document received by the R&D office is marked as Version 1, 6th October 2010.

Yours sincerely

Jonathan Cook
**Director of Corporate Services**
**NHS Norfolk**

cc: Sue Steel, University of East Anglia, Sponsor Representative
Peter Langdon, University of East Anglia, Academic Supervisor
File

Enc

**Appendix I - Version One of the MET-ID**

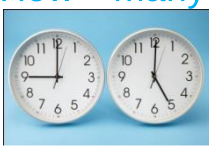
## <u>Tasks</u>

**Do the green, blue and red tasks:**

1) <u>Do these 3 things:</u>

- Find a book and give it to Tom 
-  Ask for an envelope from reception and give it to Tom when you have finished
-  Find a pencil and give it to a member of staff

2) <u>Tell Tom the answers to these questions:</u>

- How many clocks are there at the day centre? 
-  What is one of today's activities?
- 

What is today's date and month?

3) <u>Tell Tom when 5 minutes has passed on the watch</u>

## TELL TOM WHEN YOU HAVE FINSISHED

## Rules

- Do **ALL** the tasks, but in **ANY** order

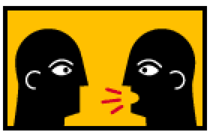
- **DON'T** go back into a room you've already been in

- **DON'T** leave the **DAY CENTRE**

- **DON'T** enter any **STAFF OFFICES**

- **DON'T** carry more than **ONE THING**

- **DON'T RUSH**



- **ONLY** speak to Tom, if it's for one of the **TASKS**

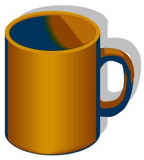
**Appendix J - Version Two of the MET-ID**

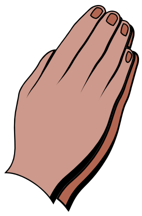
## Tasks

**Do the green tasks:**

- Find a book and give it to Tom. 

- Ask for a piece of paper from reception and put it on a chair.

- Find an empty cup and give it to Tom. 



- Clap your hands together 3 minutes after you start.

- Find a pencil and put it on a table.

**TELL TOM WHEN YOU HAVE FINSISHED**

# Rules

**Follow the red rules:**

- Do **ALL** the tasks in **ANY** order.

- **DON'T** speak to Tom unless it's to tell him when you've finished.

- **DON'T** walk inside the **RECEPTION.**

- **DON'T** walk inside any **STAFF OFFICES.**

- **DON'T** go back into a room you've already been in.

- **DON'T** leave the **DAY CENTRE.**

**Appendix K - MET- ID Administration Instructions**

Ensure that there is paper at reception and that books, cups and pencils are freely available at the day centre.

In the testing room, give the participant the clipboard with the attached exercise sheet, recorder pen and (if the participant does not have one already) a watch. Then read the following instructions to the participant:

*In this next task we will be walking around the day centre and I want you to do these green tasks. I will tell you about these green tasks, but you can also hear them read out by touching the sticker next the words with this pen* (show participant pen and how to activate the instructions in auditory form).

*The green tasks are to do the things listed here* (examiner to play the rule with the recorder pen and then describe the rule as below).

*Find a book and give it to me.*

*Ask for a piece of paper from reception and put it on a chair.*

*Find an empty cup and give it to me.*

*Clap your hands together three minutes after you start. You have a watch to help you with this.*

*Find a pencil and put it on a table.*

*Finally, tell me when you have finished.*

*However, while do these green tasks you must obey these red rules* (examiner to play the rules with recorder pen and then describe the rules as below).

*You must carry out all the tasks but you may do so in any order.*

*This means you shouldn't speak to me unless you are telling me when you have finished.*

*You must not walk inside the reception. You need to speak to someone from reception for one of the tasks, but don't walk inside it.*

*You must not walk inside any staff offices.*

*Don't go back into a room you have already been in. This means that if you have been into a room, you cannot go back into it again. However, you can go into the ... and the... as many times as you like* (depending upon the design of the day centre there may be certain areas where this is necessary to such as the foyer, hall by front door or corridors etc).

*Don't leave the day centre. This means you must not leave by any of the entrances or exits.*

*I now want to give you to try and learn these rules. I will then ask you to tell me as many as you can. Don't worry it's not a test. I just want to make sure that you know the rules. You can use the pen to help you. I will help you remember them afterwards if you get stuck.* (Make sure the participant plays through all the rules at least twice. Record the number of rules recalled immediately on the sheet. When the participant cannot remember any more, prompt the unrecalled rules until the participant is familiar with them).

**Practice:**

Move to the area by the front door of the day centre. Give the participant the practice exercise sheet and rules.

*Let's do a practice. Look at these tasks here. Can you get the pen to read them out?*

*OK, so now you must do what it says, so find a chair and sit on it*. Give participant help as necessary and point out any rule breaks.

*Find a window and touch it.* Give participant help as necessary and point out any rule breaks.

Repeat these steps once more giving appropriate encouragement and reinforcement for correct responses.

*OK so now you know what to do.* Remove practice sheet and reveal exercise sheet. *Now you need to do these tasks but not break any of the rules.* (Clarify that participant understands what they must do and go over tasks and rules again if necessary).

*During the exercise I will be following you and watching what you are doing. I can't give you any help from now onwards so please do not speak to me unless you are telling me when you have finished.*

*Begin the exercise* (Assessor to start timing at this point)

If participant asks a question during the exercise state *remember one of the rules is not to speak to me, but you must to do all these tasks* (point at board) *and not break the rules* (point at board).

If at any point during the test the participant continually scrolls through the tasks/rules with the recorder pen more than two times **without actually initiating any of the tasks** state, *remember one of the rules is not to speak to me, but you must to do all the tasks* (point at board) *and not break the rules* (point at board).

**Discontinue rule:**

When the participant says "I've finished" the test is completed, irrespective of how many/few of the tasks they have done.

If the participant does not say "I've finished", but is clearly not going to do any (more) of the tasks, use the prompt. If the participant still fails to initiate any (more) tasks ask *have you finished?* If the participant says "yes" end the task

## Appendix L - MET-ID Rule Recall Response Sheet

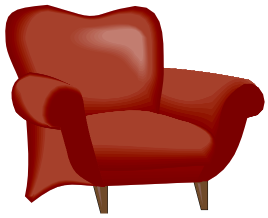
Participant Code: Date of Testing:

| RULE | Recall Yes/No | Cued Rule | Recall Yes/No |
|---|---|---|---|
| Do all the tasks, but in any order. | | *There was a rule about* **Carrying out the tasks** *Do you remember what the rule was?* | |
| Don’t speak to Tom unless it’s to tell him when you’ve finished. | | *There was a rule about* **Speaking to Tom** *Do you remember what the rule was?* | |
| Don’t walk inside the reception | | *There was a rule about* **The reception** *Do you remember what the rule was?* | |
| Don’t walk inside any staff offices. | | *There was a rule about* **Staff offices** *Do you remember what the rule was?* | |
| Don’t go back into a room you’ve already been in. | | *There was a rule about* **Going back into a room you have already been in** *Do you remember what the rule was?* | |
| Don’t leave the day centre. | | *There was a rule about* **Leaving the day centre** *Do you remember what the rule was?* | |

**Appendix M - MET-ID Practice Tasks**

## Practice tasks

**Do the green tasks:**

- Find a chair and sit on it. 

- Find a window and touch it. 

## Appendix N - Written Record of Participant Performance

22

08.04.2011

Practice tasks completed fine

| | |
|---|---|
| 00:00 | Reading tasks in hall |
| 00:22 | Walks over to shelf in hall. Gives examiner leaflet. |
| 01:05 | Walks over to reception. Walks inside reception. Appears to believe that she is not allowed to talk to staff. Whispers "paper" to staff. |
| | Staff member gives paper to participant. |
| | Participant puts paper on chair in reception. |
| 02:29 | Walks into kitchen. |
| | Gives examiner and empty cup. |
| 03:09 | Claps hands together. |
| 03:44 | Walks into recreation area. |
| 04:07 | Walking around recreation area. |
| 05:14 | Say's "finished" |

## Appendix O - MET-ID Scoring Sheet

Participant ID: Date of testing: Rater:

| **Effort category** | **Example** | **Yes?** |
|---|---|---|
| ***Task attempts***<br>*-Where a task is attempted but not completed satisfactorily.*<br>*- If the task is also completed satisfactorily then a point is still awarded for the attempt.*<br>*- Award one point per task attempt.* | • Found a book (award point even if participant finds alternate relevant item e.g., magazine, leaflet, brochure, newspaper etc). | |
| | • Asked for a piece of paper from reception (award point even if participant fails to put item on chair). | |
| | • Found an empty cup (award point even if participant finds a cup but fails to give it to examiner, if the cup has liquid in or if alternate relevant item e.g., glass). | |
| | • Clapped hands together (award point even if clap is too early [<two minutes] or too late [>four minutes]). | |
| | • Found a pencil (award point even if participant uses a pen/marker or gets pencil but fails to put it on a table). | |
| | • Told examiner when finished (award point even if participant indicates they have finished without prompting but fails to explicitly say “finished”). | |
| *Total task attempts* | | |

| Error category | Example | Yes? |
|---|---|---|
| ***Rule breaks*** *-Where a specific rule was broken.* *- Award one point per rule break.* | • Failed to **attempt** all six tasks (award if total task attempts score is five or less) | |
| | • Spoke to examiner other than to say when they had finished (includes reading out tasks) | |
| | • Entered reception | |
| | • Entered staff office (not reception) | |
| | • Re-entered **side room** (re-entering the hall/foyer and/or recreation area is OK) | |
| | • Exited day centre | |
| *Total Rule Breaks* | | |

| **Effort category** | **Example** | **Yes?** |
|---|---|---|
| ***Task completions***<br>*-Where a task is attempted and completed satisfactorily.*<br>*- Award one point per task completion.* | • Found book **and** gave it to examiner (not completed satisfactorily if item was not a book e.g., item was a magazine or brochure or does not give it to examiner) | |
| | • Got a piece of paper from reception **and** put it on a chair (not completed satisfactorily if paper was not put on a chair or put a table/floor etc) | |
| | • Found an empty cup **and** gave it to examiner (not completed satisfactorily if cup is not given to examiner, has liquid in it or relevant item given is not a cup e.g., a glass). | |
| | • Clapped hands three minutes after starting (not completed satisfactorily if participant claps hand too early [<two minutes] or too late [>four minutes]). | |
| | • Found a pencil **and** put it on a table (not completed satisfactorily if participant uses a pen/marker rather than a pencil or puts it somewhere other than a table). | |
| | • Told examiner when finished (not completed satisfactorily if states something like “I’m ready to go back to the room now” rather than specifically using the word “finished”). | |
| *Total task completions* | | |

## Appendix P - Skewness and Kurtosis for each Measure

| | | MET-ID task attempts | MET-ID rule breaks | MET-ID task completions |
|---|---|---|---|---|
| N | Valid | 40 | 40 | 40 |
| | Missing | 0 | 0 | 0 |
| Skewness | | -.829 | -.052 | .204 |
| Std. Error of Skewness | | .374 | .374 | .374 |
| Kurtosis | | -.740 | -.736 | -1.159 |
| Std. Error of Kurtosis | | .733 | .733 | .733 |

| | | DEX Total | SPT total score | TOLT total points score | WIAT word reading raw score |
|---|---|---|---|---|---|
| N | Valid | 40 | 40 | 40 | 40 |
| | Missing | 0 | 0 | 0 | 0 |
| Skewness | | .660 | .193 | -.154 | .058 |
| Std. Error of Skewness | | .374 | .374 | .374 | .374 |
| Kurtosis | | .747 | -.793 | -1.377 | -1.107 |
| Std. Error of Kurtosis | | .733 | .733 | .733 | .733 |

| | | WASI FSIQ | BPVS raw score |
|---|---|---|---|
| N | Valid | 40 | 40 |
| | Missing | 0 | 0 |
| Skewness | | 1.132 | -.299 |
| Std. Error of Skewness | | .374 | .374 |
| Kurtosis | | .583 | -.295 |
| Std. Error of Kurtosis | | .733 | .733 |

# Appendix R - Kolmogorov-Smirnov and Shapiro-Wilk tests for each Measure

**Tests of Normality**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| MET-ID task attempts | .248 | 40 | .000 | .826 | 40 | .000 |
| MET-ID rule breaks | .225 | 40 | .000 | .849 | 40 | .000 |
| MET-ID task completions | .193 | 40 | .001 | .913 | 40 | .005 |
| DEX Total | .109 | 40 | .200* | .961 | 40 | .185 |
| SPT total score | .114 | 40 | .200* | .964 | 40 | .228 |
| TOLT total points score | .191 | 40 | .001 | .914 | 40 | .005 |
| WIAT word reading raw score | .114 | 40 | .200* | .955 | 40 | .111 |
| BPVS raw score | .105 | 40 | .200* | .965 | 40 | .254 |
| WASI FSIQ | .198 | 40 | .000 | .883 | 40 | .001 |

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.