Computer-based Assessment of the Executive Functions in Children

Matthew Howell Jones Chesters

University College London

A Dissertation Submitted for the Degree of

Doctor of Clinical Psychology

August 1998

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Recent work in experimental psychopathology has shown that the profound developmental psychopathologies, such as autism and ADHD, are associated with deficits in frontal lobe or executive functions. These results may afford new measures of psychopathology, and help in the modelling of disorders, or generation of interventions. However, different disorders are associated with different executive deficits; the full range of executive functions have yet to be investigated; and there may be quantitative rather than qualitative differences between disordered and normally developing children. This study reports on the development of a computer-based battery for assessment of the executive functions in children, and provides preliminary data on its use. Normally developing children (five boys and five girls, from each of the school years 3, 4, 5 and 6) were assessed on five executive function tests. One test assessed cognitive inhibition; another motor inhibition. Three tests involved shifting mental set from one task to another (colour- or shape-naming). In the simple shift condition, colour patches or shapes appeared, and the child had to shift tasks every two trials. In a cross-talk condition, coloured shapes were presented (which could be colour- or shape-named) but the test still involved predictably shifting tasks every two trials. In a cued shift condition, colours and shapes were presented unpredictably, so that only the stimulus cued the response. The results show that children's speed and accuracy on most tests tends to improve with age, but performance was relatively independent of general intellectual ability, and response time is a better index of performance than error rates. The implications and opportunities for future research using similar paradigms are discussed.

Acknowledgements

Special thanks to Olly Stephens for help writing the computer program CHROME used in this research, and for his patience over the years. Many thanks to Mr Stephen Lancashire and his staff at the Canonbury Road School, Islington, for cheerfully recruiting pupils to the study. Thanks to Dr Tony Charman, my supervisor, for advice and encouragement: the same to Emma Brown for help co-ordinating the project. I am very grateful to my sister, Rebecca, for her enduring support.

This thesis is dedicated to the memory of my Taid, Howell Jones

> "The capacity for control (sophrosýnê), the ability to dominate oneself, to govern things, the sharpness of the eye, the sober choice of the means to achieve an end: all these things detach the mind from those powers that came before Athena."
>
> Roberto Calasso, *The Marriage of Cadmus and Harmony*.

Table of Contents

Introduction

Since James (1890) devoted a chapter of his Principles of Psychology to "The Will", the control of mental processes and action have fascinated, and been an enigma to, psychologists. With the advent of the information processing metaphor in cognitive psychology, theoretical models of a controlling or "executive" mechanism have been developed (e.g., Supervisory Attentional System, Norman & Shallice, 1986; Working Memory, Baddeley, 1986). Also, failures of control in cognition and action have been comprehensively described: both in normal populations (e.g., Reason, 1990) and in patients with brain damage (e.g., Shallice & Burgess, 1991).

Evidence from brain-damaged individuals suggests that the failure of control is intimately associated with lesions to the frontal lobes or pre-frontal cortices (Duncan, 1995). The discovery of this relationship has led to the development of the "frontal metaphor" that presently underlies the construct of executive functions in cognitive and experimental neuropsychology (Pennington & Ozonoff, 1996). Thus, the attributes of the executive have been definable both "theoretically" in cognitive psychology, and "empirically" in the neuropsychological investigation of individuals with frontal lesions.

1) Executive Function and Experimental Psychology

a) Background

The term "executive function" is a relatively recent one in experimental psychology, and its appearance is directly linked to the shift towards the cognitive or "information processing" paradigm in psychology since the 1960s. One issue is this: if the functions or elements of the cognitive system (perception, attention, learning, memory, language etc.,) are modular (Fodor, 1983), how can they act in concert to generate meaningful processing and behaviour without control systems? Newell (1980), a critic of the "executive" construct, has concisely summarised its early history, and refers to it as the "homunculus": a little man in the head who does important processing tasks, and presumably has a similar little man in his head, and so on. "A major item on the agenda of cognitive psychology is to banish the homunculus … it is the homunculus that actually performs the control processes in Atkinson & Shiffrin's famous memory model … does all the controlled processing in the more recent proposal of Shiffrin & Schneider … who is renamed the executive in many models, clearly a promotion," (Newell, 1980). Theories of executive function are therefore related to specific models of other aspects of cognitive functioning.

In the main, executive function models have emerged from theories of attentional selection, the structure of memory, and the relationship between attention and learning or memory. As will be shown, the nature of executive functions, indeed their existence at all, is very much a matter of debate. Most models rely on evidence from different aspects of dual task performance i.e., where a participant is performing two simple or complicated tasks concurrently or in quick succession. More recent models have also depended to a large extent on the data available from cognitive neuropsychology and the investigation of patients with brain damage to the frontal lobes. This is particularly true for the “specialist controller” approaches (outlined below), and to a large extent has motivated the advances that have been made in “connectionist” modelling of control functions. Here, the theoretical aspects of these models will be considered in some detail, while the neuropsychological evidence will be examined in the next section.

b) General Purpose Central Processor Approaches.

General purpose central processor approaches locate the putative properties of executive function in the limited capacity of a “single channel” or “central processor” system. A core assumption in attention research has been that the limited information-processing capacity of the brain imposes a selective character on cognitive operations; the function of attention is to protect this single channel system from information overload. This assumption underlies Broadbent's (1958, 1982) model of attention: the registration of simple physical qualities of sensory input can occur rapidly in parallel, but the semantic categorisation or recognition of objects can not. Sensory input therefore enters a limited capacity general purpose central processor only after passing through a "bottleneck" where information is selected or attenuated in signal strength, according to whether it contains certain physical features. The limited capacity of this central processor is responsible for the difficulties participants have when performing two tasks simultaneously (e.g., shadowing an auditory message while copying text). In such situations, compared to performing either task by itself, less capacity of the processor is available for each task, so performance suffers.

There is no executive function in this system, only an emergent property of "selection" in the central processor mechanism. Of course, general-purpose central processor theory has difficulty in accounting for cases where dual task interference is minimal. Allport, Antonis and Reynolds (1972), for example, found little dual task decrement (relative to the single task condition) when experienced pianists sight-read music while concurrently shadowing an auditory message. But Broadbent (1982) noted that the prose and music "messages" which are processed in such tasks are not only highly structured (and therefore to some extent predictable) but also there is the opportunity to preview the message before responding. Thus participants may be able to "switch" the general purpose central processor from one message to the other.

A more recent approach based on the "single channel" model has been advanced by Pashler (1993) who utilises the "psychological refractory period" (PRP) paradigm. In the classic PRP experiment two stimuli are presented, separated by a short interval (stimulus onset asynchrony, SOA), and each requiring a separate speeded response. As the SOA is reduced a "critical point" is reached beyond which there is a slowing of response to the second stimulus: the PRP effect. In some cases the latency effect is exactly equal to the difference between SOA and the "critical point", suggesting that there is a bottleneck in the attentional system at which some or all processing of the second stimulus is postponed until some or all processing of the first stimulus is complete. In his review of a number of studies on which processing stage might be involved, Pashler shows that an explanation in terms of a response-postponement process is appropriate regardless of the modality of response output.

To simplify, stimulus 2 may not cross the bottleneck until the response to stimulus 1 has gone through: if stimulus 2 arrives a while after stimulus 1, it enters a response selection mechanism (RSM) immediately and there is no PRP effect; if it arrives early (before, at the same time, or too soon after stimulus 1) it must wait and there is PRP latency. Pashler has confirmed this interpretation by showing that if the perception of stimulus 2 is delayed the PRP effect is abolished. Thus, no executive functions need to be specified: participants can "set themselves" to perform two tasks at once as well as they can a single task, and there is task interference only when responses arrive at the processing bottleneck. One problem with Pashler's model is that he has identified a number of "exceptions to the rule". The postponement effect seems not to apply to (voluntary) saccadic eye-movements, or tasks involving well established or highly compatible stimulus-response associations (e.g., shadowing verbal input).

Allport (1993) has roundly criticised central processor approaches for neglecting the contribution of control processes, and indeed for implicitly assuming them. Precisely what, one might ask, achieves the "switching" between messages that Broadbent assumes to occur during some dual task performance? Furthermore, in Pashler's architecturally well developed model, what is this rather ambiguous response-selection mechanism which is not formulated in any detail? Allport (1993) asks, "Is the bottleneck to be thought of as the *cause* of attentional selectivity ... or be understood as the *consequence* of other co-ordinating processes involved in maintaining the coherence of purposive action?" Allport argues that executive processes are needed to maintain engagement to particular cognitive tasks, to facilitate memory search, and to protect ongoing behaviour from interference by other demands or sources of information.

Allport is not alone in recommending that one (or many) specialist executive function mechanisms exist relatively separately from other domain-specific processes (such as those involved in perception, stimulus classification or response execution). As noted by Newell (1980), the distinction between control and other cognitive processes was initiated in theories of the functional architecture of memory (Atkinson & Shiffrin, 1968) and the control of attention and performance (Shiffrin & Schneider, 1977). These ideas have been developed in terms of memory systems by Baddeley (1986, 1990) and for control of action by Shallice (1988; Norman & Shallice, 1986; Shallice & Burgess, 1991).

c) Specialist Controller Theories

i) Memory

On the basis of evidence from several memory tasks, Atkinson and Shiffrin (1968) proposed that memory should be viewed as comprising a separate short term store (STM) and long term store (LTM): the Two Process model of memory. Limited information from the sensory modalities entered STM where it could be held for a limited period for "coding". Coding was achieved by a set of *control processes*, one of which was the "rehearsal buffer" which served to maintain information in STM and to transfer information into LTM (or otherwise it was forgotten). The control processes involved (including the nature of rehearsal) were not specified in detail, but were deemed necessary for the maintenance, elaboration and transfer of material within and between the two stores. The Two Process model had obvious problems, but was significantly revised by Baddeley (1986, 1990) who distinguished several sub-systems within STM and termed the construct Working Memory.

Working Memory may be defined as a system for the maintenance and manipulation of information in the short term. The system comprises three sub-systems. The central executive (CE) is an attentional control system which co-ordinates the operation of two slave systems, the phonological loop (which retains verbal material) and the visuo-spatial scratch-pad (which retains visual material) all of which have component input and output elements. In his original formulation, Baddeley (1986) mapped the CE directly onto Norman and Shallice's (1986) early conception of the supervisory attentional system (SAS, discussed later), but more recently the existence and properties of the CE have been investigated directly in experiments on dual task performance involving the putative slave systems.

Baddeley, Della Salla, Gray, Papagno and Spinnler (1997) summarise a number of studies which used a combination of digit span rehearsal (phonological loop material) and a tracking task (visuo-spatial scratch-pad material) in a dual task paradigm. Patients with "dysexecutive syndrome", either as a result of Alzheimer-type dementia or focal frontal lesion showed substantially greater decrement in the dual task conditions than matched intact normals, even when there was no significant impairment in the single task conditions. But normals, too show a noticeable performance decrement on task combinations of this kind, and Baddeley et al., conclude that their index of dual task performance impairment maps squarely onto the efficiency of a central executive system.

ii) Attention and Performance

One of the most compelling phenomenon noticed in experimental paradigms, and in everyday life, is the special advantage brought to task performance by practice. While many tasks and activities are initially slow, error-prone and require concentration, some even very complicated activities become effortlessly fast and accurate with repeated experience (e.g., reading music, driving a car). The idea that some tasks run off "automatically" while others have to be "controlled" suggests distinct levels of task performance. Shiffrin and Schneider (1977; Schneider & Shiffrin, 1977) outlined a very influential model of attention and automaticity. They saw automatic processes as qualitatively distinct from controlled processes: automatic processing is independent of resources, involuntary, inflexible, fast and parallel; controlled processing is resource-dependent, voluntary, flexibly employed, slow and serial. Moreover, only certain forms of practice, where the same stimuli are consistently associated with the same response, will allow a process to become automatic.

Shiffrin and Schneider demonstrated, using display and memory scanning tasks, that consistent stimulus-response mapping (CM; e.g., digits are always targets, letters are always distractors) ultimately produced fast, accurate performance that was unaffected by memory load. Performance on varied mapping (VM; targets are sometimes digits, sometimes letters) did not improve in the same way. Also, if the CM task was reversed after extensive practice (i.e., now letters are always targets, and digits are distractors) performance was massively impaired suggesting that CM mapping produces involuntary and inflexible processing.

However, the automatic-controlled distinction, and Shiffrin and Schneider's model in particular, has been widely criticised for being dichotomous rather than dimensional, and descriptive rather than explanatory. The model does not elaborate the architecture or memorial changes that take place to make a process automatic (Logan, 1988), and does not allow a "continuum" between automaticity and control. The contribution of different processing components may vary, and different degrees of automaticity might be achieved. Nevertheless, the distinction between controlled and automatic processing has provided a useful introductory metaphor for the related distinction between "executive" and "non-executive" functions. Theorists who have focused more on the strategic aspects of processing have emphasised the likelihood that there are multiple levels of control.

Shallice (1988, Norman & Shallice, 1986; Shallice & Burgess, 1991) proposes a model of control based on the "production system" architecture developed in computer science (Allport, 1980). A production system architecture is one which comprises a vast set of intensely task-specific condition-action units, termed "productions", and a "work space" which receives information externally from the sensory apparatus and internally from the operation of productions. Each production continually monitors the workspace until it identifies information that matches its condition. When it does, the production performs its operation and places the results back into the workspace: the system is driven only by the procedures available in the productions, an information available in the workspace. In Shallice's model, there are three levels of control. At the lowest level, "productions" are thought and action *schemata*: non-conscious processes which run off fully automatically in response to specific triggers. Schemata are organised by experience into hierarchies: with practice, or when skilled, a portmanteau schema triggers its subordinate component schemata to produce rapid, accurate operations. At the next level, partially automatic *contention scheduling* operates to resolve conflict between activated schemata, and impose order on the anarchic triggering of all appropriate schemata in response to a given stimulus. Contention scheduling involves a process of mutual inhibition between schemata such that only the most strongly activated manages to gain, and keep, control of behaviour.

However, a system using only salience, frequency or recency to choose between competing schemata would be a prisoner of habit. Thus, at the highest level an executive mechanism, the *supervisory attentional system* (SAS) serves to deal with instances where automatic task activation would not be adequate. Shallice and Burgess (1991) identify five such situations: where the task is novel or not well-learned; where the task involves planning or decision-making; where the task is difficult or dangerous; where there is pressure for error correction or trouble-shooting; and where a strong, recently primed or habitual response has to be overcome. The SAS has access to the external environment, the individual's state, task demands and intentions, and modifies lower level operations accordingly by biasing the contention scheduling mechanism. The clear strength of this model is in moving beyond the automatic-controlled distinction (Monsell, 1995). The automatic content-based aspects of cognition are neatly captured in the production system architecture of the model, while the supervision of attention at other times can be viewed as a “biasing” or “modulation” of the automatic elements in the system, rather than the seizing of control by an executive overlord.

The main problem for this model (indeed most production system models) is that it does not explain how the system can be configured to perform a novel task, but this is one of the key conditions Shallice outlines requiring intervention by the SAS. Obviously, people are required to deal with novelty a good deal of the time.

Participants in psychology experiments, for example, perform novel tasks (like visual probe detection or lexical decision) very fluidly with a little practice. However, the appropriate schemata are not already in the processing architecture, and have to be configured. The question is how are the control operations of stimulus input and response output linked in the initial absence of an appropriate task schema? Presumably, the SAS performs this function, but how it achieves this is not specified.

The CE in Baddeley's Working Memory, and the SAS in Shallice's production system model, might be the same entity. Even if the are not, there are inherent problems for these theorists in proposing monolithic, central, controller systems of this kind. First, if it does not prove possible to deconstruct them, then they may be indistinguishable from the kind of limited capacity, general purpose, central processors they were intended to replace, and accordingly they begin to resemble the processing homunculus of the classic critique. Again, Allport (1993) has been critical of the central "controller" approach. He argues that the evidence from experimental psychology and neuropsychology demands that the CE or SAS should be disassembled into several disparate executive function systems (leaving no central element behind) and that control mechanisms are heterogeneous and distributed.

d) Connectionist Approaches

The idea that there are multiple, separate control systems distributed across the cognitive architecture fits well with connectionist approaches to executive functions. The key theme in connectionism is that information processing occurs in parallel and is distributed over the cognitive architecture. One early attempt to approach attentional control in these connectionist terms was outlined by Cohen, Dunbar and McClelland (1990). They modelled performance on the Stroop task in which the participant is required to name the ink-colour in which a word is presented. The classic effect is that it takes longer to name the ink-colour when the stimulus is an incongruent colour-word (e.g., "RED" printed in green ink) than when the stimulus is meaningless (e.g., "XXXX"). Control functions are assumed to bias responses in favour of an ink-naming network despite interference from a word-reading network. Cohen et al., (1990). modelled two feed-forward networks: one mediating performance of word naming, the other mediating performance of colour naming. Each network has a layer of input units and output units, between which there is a layer of "hidden units". The activity level of these hidden layers is modulated by a single, independent "task unit". The task units specifically and locally represents the stimulus-response mapping of each network. Thus the relative activities of the two task units determine which network is dominant and whether response output will be the ink colour or the word name.

This approach has been extended by Kimberg and Farrah (1993) who modelled four tasks which have been assumed to depend heavily on executive function, including the Stroop task and the Wisconsin Card Sorting Test (discussed later). Kimberg and Farrah constructed separate and independent models for each of these four tasks using a production system architecture. Each task was managed by a single, dedicated "information-specific working memory system", the interaction of elements in which could over-ride the priming activation of a recently operated task set. The simulation's performance on each task was compared across conditions where connection strengths between the elements of working memory were either maximally weighted or degraded. They found that low-level degradation of weights in each system produced performance limitations directly comparable to those errors (particularly perseveration errors) observed in participants in the cognitive laboratory. Thus Kimberg and Farrah have shown that logically identical working memory "procedures" can control task performance in very different information-specific systems. Their results suggest that there is no need to posit any general "executive system", rather there may be any number of task-specific working memory systems in the cognitive architecture, and control is a distributed memorial function.

However, Kimberg and Farrah (like Cohen et al.,) have modelled performance and performance impairment in systems that are assumed to exist already, and where task representations have developed to a stable state. They do not offer information on how novel tasks and associated working memory systems become so structured. Schneider (1993; Schneider & Detweiler, 1988) has developed a connectionist approach to studying the changes involved in the transfer from "controlled" to "automatic" task performance (outlined above) that might contribute to this issue. Schneider and Detweiler (1988) propose that learning to perform even a simple task individually involves moving through a number of stages. Automatic and controlled processing are considered the end points on a continuum of performance efficiency. When a novel task is attempted, it is highly controlled and makes great demands on working memory to co-ordinate possible input and output representations. With practice, a stage of "context controlled comparison" is reached, where a limited set of the necessary memory vectors are responsible for the representation of task information. With further practice, a stage of "goal state maintained" controlled comparison is reached, where memory vectors are transmitted from input to output directly and very efficiently. Eventually, full automaticity is achieved, where there is very little draw on memory load and processing is highly efficient.

Automatic dual-task performance entails several other changes, including increased use of non-competing resources, increased parallel processing, and increased "chunking" of information in memory transmissions. The Schneider and Detweiler (1988) model maps very naturally onto a production system architecture, and implies the development of "higher order" controllers dedicated to each task-specific system (Rabbit, 1997). Schneider (1993) has described how different controllers may exist for different types of *information*, and that controllers may themselves be organised into hierarchies of control with second or third order controllers overseeing the co-ordination of several task-specific systems (e.g., in dual task performance). This is an important idea, since it moves us beyond the difficulties of both the monolithic "central executive" and the distributed but potentially uncoordinated task-specific controllers. It clearly represents an advance on Shiffrin and Schneider's (1977) automaticity theory, but is compatible, for example, with their observation that dual task performance on even difficult tasks can be very efficient provided the two tasks are practised together.

The remaining issue for connectionist models is to define in detail the nature of their "task units", "task-specific working memory systems" or "memory vectors". In the models outlined above, the implicit suggestion is that the (many) control structures involved are themselves productions of some kind: they respond to detected information and activate the relevant component networks. In the Cohen et al., model, task control units are primarily "excitatory": they engage, facilitate or enhance processing in task-specific systems. However, in the Kimberg and Farrah model, the elements of working memory are "inhibitory": they suppress activity in a task system which would contribute to processing a now-inappropriate task. The idea that the inhibitory control of task-specific systems is a crucial property of executive functioning which can be represented in a connectionist architecture is a significant advance. Inhibition is also a feature of the lower levels of control in Shallice's model (i.e., contention scheduling) and is compatible with one of the main themes to emerge from neuropsychological investigation: that the frontal cortex provides an inhibitory influence on other areas of cortex (Denny-Brown, 1958). It is to this arena that we now turn.

2) Executive Function and Neuropsychology

a) Background

As Lezak (1995) notes, the important role of the pre-frontal cortex in cognition and action has long been recognised by neuropsychologists. First, the frontal lobes were observed to have developed (to a substantial size) most recently in human evolution, and early researchers thematically associated the pre-frontal cortices with the highest cognitive functions. Second, evidence from human patients who had suffered trauma to the frontal lobes, and from lesion experiments on the equivalent regions in animals, suggested that the area was intimately involved the organisation and planning of purposive behaviour. Nevertheless Hebb (1945), in his substantial review of the evidence available at that time and his own research on the "intellectual abilities", was led to conclude that, "no one has proved that any single form of normal behaviour is dependent on this part of the brain."

Contemporary interpretation of the neuropsychological evidence is that the pre-frontal cortices are not responsible for the representation of specific skills, information, or low-level abilities, and thus damage to the frontal lobes need not be evident from neuropsychological assessment involving single step tasks, highly directed behaviour, and closed or concrete questions (Lezak, 1995). Rather, the pre-frontal cortices are involved in "orchestrating" interactions between the main functional systems (e.g., perception, memory, attention) and the affective, motivational and motor-output systems. Damage to this region therefore will be most evident in problems with daily living and certain behavioural disorders. Lezak divides these behavioural disorders into five main groups: problems with starting (e.g., lack of initiation or spontaneity), problems with stopping (e.g., impulsivity, disinhibition), problems in making mental or behavioural shifts (e.g., response perseveration, stereotypy), deficient self-awareness (lack of monitoring or correction) and concrete attitude (literal-mindedness, failure of planning).

From this perspective, the relationship between the frontal cortices and the executive functions becomes more transparent. Clearly, therefore there are a number of highly complex functions that the executive must perform, including maintaining or shifting mental set, controlling interference or response competition, integration of spatial and temporal information, monitoring, inhibition and planning. Accordingly, there are sets of cognitive tasks which have attracted both theoretical interpretation or modelling in terms of executive function, and having been shown to be affected by damage to the frontal brain regions are utilised as tests of neuropsychological functioning. I will review these in turn. However, the assessment of executive functions is by no means an uncontroversial business. As I shall show, the reliability and validity of not only the executive function tests themselves, but also the psychological constructs on which they depend, and their relationship to the pre-frontal cortices, has proved a matter of considerable debate.

b) Tests of Executive Functioning

The five related groups of cognitive task most frequently employed in neuropsychology are the planning, working memory, fluency, set-shifting, and inhibition tasks.

i) Planning Tasks

Planning tasks typically require the participant to decide upon and order the series of steps required to reach a goal state, and to execute them. The hallmark planning task is the Tower of London test (e.g., Shallice, 1988) in which the participant is presented with a large, medium and small ring, placed in size-order to form a "tower" on the first of three rods. The task is to move the rings, one at a time and never placing a larger one on a smaller one, onto the other rods, eventually rebuilding the tower on any rod except the first. Executive functions are involved in configuring a representation of the goal state and intervening stages, monitoring progress and dealing with errors. Shallice and Burgess (1991) have shown that patients with frontal lobe damage take longer than normal controls to rebuild the tower, and do so in many more moves. Other planning tasks include Maze Tracing procedures (e.g., Porteus, 1965) which patients with frontal damage have been shown to solve more slowly than either matched controls or patients with either posterior damage or severe head injury (Levin, Goldstein, Williams & Eisenberg, 1991).

ii) Working Memory Tasks

Working memory is a construct comprising short-term stores for verbal and visuo-spatial information, and a controlling executive. Working memory tasks are therefore those that require the participant to organise sources of information and perform operations on them. For example, the Cognitive Estimates Test (Shallice & Evans, 1978) is a set of brain-teasing questions such as, “How fast does a racehorse run?” or “How many elephants are there in London?” the answers to which are not directly given by common knowledge. The answers are arrived at instead by summoning up diverse sets of information and operating on them in a logical way to produce a valid inference (e.g., “Are there any elephants in London? Where? Zoos? How many zoos in London? How many elephants in each?). Shallice and Evans (1978) demonstrated that patients with frontal lobe damage often gave very inaccurate (and sometimes frankly bizarre) answers to questions on this test. Other Working Memory tests include mental arithmetic tasks such as the Arithmetic sub-test of the Wechsler Adult Intelligence Scales (e.g., WAIS-R; Wechsler, 1981). Although performance on such tasks tend to depend more on the efficiency of short-term stores than the executive per se, Arithmetic has been shown to load on the “Freedom From Distractibility” factor of WAIS profiles rather than on the standard Verbal or Performance domains.

iii) Verbal and Performance Fluency

Fluency tasks are those that require the participant to be highly productive in structuring response outputs according to given rules. For example, in the Controlled Oral Word Association Task (Benton & Hamsher, 1989) the participant is asked to say out loud as many words as they can think of, in one minute, beginning with a given letter of the alphabet, but not including names, numbers or stems with differing suffixes. Good performance depends on the deployment of strategies such as using phonological trends (e.g., far, for, fair, fur), or semantic themes (e.g., salt, sour, sugar, sweet). Executive functions are required to co-ordinate strategy use, monitor repetitions, and inhibit invalid responses (e.g., shoe, shoes, shoe polish). (In clinical practice, it is always interesting to note how many taboo words, for example beginning with F, A or S, are produced.) Patients with frontal lesions tend to have markedly depressed verbal fluency scores. Left frontal lesions affect verbal fluency more than right-sided damage, but bilateral lesions produced the most profound impairment (Lezak, 1995). The related Design Fluency task, in which the participant is required to draw as many abstract forms as they can, shows a parallel pattern of sensitivity to right-sided frontal damage (e.g., Jones-Gotman & Milner, 1977).

iv) Set-Shifting Tasks

Set-shifting tasks require the participant to shift flexibly from one mental (or response) set to another. The hallmark set-shifting test is the Wisconsin Card Sort Test (WCST, Grant & Berg, 1948; Heaton, 1981) in which the participant is required first to respond according to one task-set or rule (e.g., "sort the cards according to colour") and then to shift sequentially to different task-sets (e.g., "sort the cards according to shape"). The rules are not made explicit; rather the participant must deduce the rule from the pattern of reinforced responses (guess). In addition to a measure of overall task success, scoring the WCST provides information on specific areas of executive dysfunction, including failure to maintain an appropriate set, failure to shift to correct set, and perseveration errors. Since Milner (1964) demonstrated that patients with frontal lesions are profoundly impaired on all dimensions of the WCST, the test has been repeatedly refined, modified and standardized for use as a neuropsychological instrument (Heaton, Chelune, Talley, Kay & Curtis, 1993). Another influential set-shifting task is the elegant Trail Making Test (Armitage, 1946) in which the participant must first join-up numbered circles (in ascending order) on a piece of paper (part A) and then switch to alternately joining numbers and letters (part B). The Trail Making test, especially part B, is very sensitive to most kinds of brain damage, but Segalowitz, Unsal and Dywan (1992) have shown, using electrophysiological measures, that performance is directly related to the quality of frontal activation.

v) Inhibition Tasks

Inhibition (or interference) tasks are those which require a response to a stimulus according to one task-set despite competition from a different task-set which has to be attenuated. The hallmark interference task is the Stroop colour-naming task (Stroop, 1935; Klein, 1964) in which the participant is required to name the ink-colour in which a word is presented. The classic effect is that it takes longer to name the ink-colour when the stimulus is an incongruent colour-word (e.g., "RED" printed in green ink) than when the stimulus is meaningless (e.g., "XXXX"), but any familiar word (e.g., "HOUSE") or pronounceable non-word (e.g., "GAKE") will produce some increase in colour-naming latency. Executive functions are required to maintain responding according to the colour-naming task-set despite interference from the (well-practised) word-reading task-set (Monsell, Taylor & Murphy, 1995). Performance on the Stroop task is intimately associated with frontal lobe function (Pardo, Pardo, Janer & Reichle, 1990), and appears most profoundly depressed by left-sided frontal lesions. Other inhibition tasks include Response-Opponence tests in which the participant must give a response conventionally opposite to the one expected. For example, Drewe (1975) neatly demonstrated that patients with frontal damage were significantly slower and less accurate than normal controls to press a blue switch whenever a red light was lit, and a red switch whenever a blue light was lit.

c) Methodology of Executive Function Assessment

In his extensive review of the theory and methodology of executive function assessment, Rabbitt (1997) suggests that research in this arena is dogged by the poor reliability, uncertain validity, and low task purity of executive function measures. The issues underlying these problems are complex and highly inter-related, but they are worth considering in a little detail before turning to more recent conceptualisations of executive function assessment. I shall outline problems with the "benchmark" executive function measure, the Wisconsin Card Sorting Task (WCST) as an examplar for these difficulties.

i) Low Reliability

That measures of executive functioning have a poor test-retest reliability ("they do not predict performance on themselves") appears to be an open secret in cognitive neuropsychology (Rabbitt, 1997). First, most researchers agree that tests of executive function must be *novel* to the participant, since well-practised tasks are further along the automaticity continuum (Shallice & Burgess 1991). The existing measures certainly have a novel (in fact rather odd) feel to them, and require participants to engage in tasks not usually found in everyday life, such as naming the ink of colour words or guessing the number of elephants in London. But a task cannot be novel more than once, and is not novel for very long while it is being performed. Also, as Phillips (1997) shows, there is no guarantee that a task is novel to a given individual: crossword addicts may be at an advantage on tasks requiring memory search according to initial letters (verbal fluency), and individuals may well have prior experience of "brain-teaser" tasks like the Tower of London or Cognitive Estimates Test. Phillips also shows that the usual parallel forms approach adopted in neuropsychology (same format, different content) cannot be used in executive assessment since both the format and content of the measure must change in order for the task to be novel.

Second, tests of executive functioning tend to be difficult or unpleasant. By definition, the tasks involved demand effortful, controlled processing, and are designed to resist the development of automaticity. This is an aspect of the "oddness" and novelty of the measures, but it may mean that participants find the instructions too difficult to comprehend and the task itself too difficult to complete (e.g., on the WCST, discussed later). Phillips (1997) comments that this, "distress and demotivation might confound measurement of task performance." These additional sources of measurement error have significant implications for the validity of executive function measures, and also therefore the construct of executive function itself.

ii) Low Validity

The criterion validity of executive function measures has been examined both for predictive quality (sensitivity to pre-frontal lesions) and convergent quality (correlation with other measures of the same construct). The predictive validity of executive function measures is difficult to assess. First, the element of circularity in this process is highly problematic (Phillips, 1997): statements of the form, "if executive function tasks are those on which individuals with pre-frontal lesions perform poorly, then executive function is associated with the pre-frontal cortices, and tasks performed poorly by patients with pre-frontal lesions are executive tasks" are not very satisfactory.

Second, patients with pre-frontal lesions do not always perform poorly on (any) measures of executive functions, while patients with lesions in other brain areas often do, and there is large variability in the performance of even severely frontally-damaged patients across these measures (Reitan & Wolfson, 1994). No compelling relationship between lesion location and performance impairments has been identified, and the function-anatomy mapping between executive and frontal functions is very disappointing when compared to that observed, for example, in visual perception or language abilities.

Accordingly, the convergent validity of executive function measures, in individuals with pre-frontal lesions, has proved difficult to demonstrate. For example, in validating their Behavioural Assessment of Dysexecutive Syndrome (BADS) battery, Wilson, Evans, Alderman, Burgess and Emslie (1997) found the correlations between their measures to be acceptably high, whereas others (e.g., Kopelman, 1991; Reitan & Wolfson, 1994) have found these relationships to be weak. Burgess (1997) therefore concludes that executive functions are fractionated, task-specific, and distributed in the frontal systems. (Note that this interpretation of the evidence is compatible with the connectionist models of control outlined earlier, but goes against the specialist "central" controller approaches.)

Despite its utility then, the construct validity of "executive function assessment", is under-specified. The main problem here is that those task demands that are assumed to underscore particular tests (e.g., inhibition, interference, planning, set shifting, monitoring) themselves have low construct validity. For example, "planning" would presumably be involved in any task where some preparation to respond is made; "inhibition" may be an emergent property of successful resistance to "interference"; and "inhibition" of one mental set may well be a necessary component of successful "set shifting"; surely "monitoring" is involved in any task where errors are a possibility. In his discussion of the validity of such terms, Rabbitt (1997) warns that, "a very wide range of apparently dissimilar behaviours may all derive from the same small number of assumed system properties," and Kimberg and Farrah's connectionist model has intimated as much.

iii) Task Impurity

Most of the reliability and validity issues outlined above are also a reflection of the fact that, since executive functions are detectable chiefly in their operation on other parts of the cognitive system, then executive function measures always involve other cognitive processes. Weiskrantz (1992) noted that all neuropsychological tests are "task impure" in this way: they require the cognitive ability of interest (e.g., construction) to be adequately supported by other cognitive systems (e.g., visual perception, sensory-motor skills) which are assumed (for the moment) to be intact. But task impurity is a particular difficulty for executive function tests since by definition they involve the co-ordination of several cognitive systems and component processes simultaneously.

Performance on the Trail Making Test part B, for example, involves intact verbal comprehension (of the instructions), visual perception, letter identification, digit identification, rote alphabet recall, rote numeric recall, set shifting, short term memory, visuo-graphic skills and motor output co-ordination. Poor performance on an executive function task may therefore reflect impairment in any one or more of the many additional "supporting" systems rather than in the control process itself. A useful way to address this issue is to assess performance on a similar test without the putative executive component. For example, performance on part B of the Trail Making Test can be compared to performance on part A, which does not involve shifts of set. But this is again a partial solution, since though part A of this test does not require the letter identification and rote alphabet recall involved in part B, neither does it place as much strain on memory load or visual search. Therefore one cannot deduce executive deficits from poor performance on one (or even more) executive function measures alone; certainly a demonstration of integrity in other cognitive systems would be a necessary adjunct to an assertion of executive impairment.

In fact, Duncan (1995) has demonstrated that executive functions measures may be best validated (and predicted) by certain measures of "intelligence", and that executive functioning loads highly on only *one general factor*. Psychometric approaches to intelligence have tended to distinguish two key facets of the construct: "fluid" and "crystallised" intelligence. Crystallised intelligence reflects verbal abilities and the acquisition of cultural or general knowledge through experience (factor g_c) and tends to increase with age. Fluid intelligence reflects flexible performance on novel and complex tasks (factor g_f) and tends to diminish with age. Intelligence tests that rely less on verbal, general, or cultural knowledge, including the Raven Progressive Matrices scale (Raven, 1960) and the Culture Fair Test (Cattell & Cattell, 1960), are assumed to index g_f and Duncan (1995) has shown that scores on the Culture Fair Test strongly predict performance on his Goal-Neglect set shifting task, and other aspects of executive functioning. The suggestion is that executive functions may depend as much on one broad aspect of cognitive functioning (e.g., information processing speed or memory capacity) as on particular elements of cognitive control. Moreover, there will be individual differences in executive functioning, much as there are in g_f, that will have to be taken into account.

iv) Example: The Wisconsin Card Sort Task

To illustrate the rather thematic issues discussed above, it will be useful to consider a concrete example of the kinds of problems evident in one "frontal" or executive function test. The Wisconsin Card Sort Task (WCST) is certainly novel, complex, and effortful for participants, but in clinical practice it can be difficult to administer because of this complexity and rather aversive rubric. The apparently random rule-shifting involved in the WCST produces confusion, anxiety or frank irritation in some participants: the test can take a punishing 50 minutes to administer to very impaired adults and children. Heaton et al., (1993) examined the test-retest reliability of the WCST in children and adolescents and found highly variable generalizability coefficients, while Pennington and Ozonoff (1996) report that they have data showing that although test-retest reliability is low for WCST subscales in normal children and adolescents, it is rather higher for older samples and impaired populations. Lezak (1995) comments, "the WCST no longer measures the problem-solving ability of subjects who have solved it once … it is a "one-shot" test."

The WCST does have the brawny advantage of simultaneously measuring several dimensions of executive functioning (maintaining set, shifting set, perseveration of set) and its face validity is accordingly high. But WCST performance also depends crucially on a rather different skill: deducing rules from the pattern of reinforced responses. It is by no means theoretically or empirically given that the learning and memory components of rule deduction procedures are fundamentally executive or control functions. In addition, to this brand of "task impurity", the WCST must be supported by intact cognitive systems for processing colour, form and number, and pattern matching, as well as the visual input and motor output systems. Moreover, the WCST produces measures based only on error proportions and trials-to-criterion totals. Since response latencies and completion times are not collected, important information on processing speed and resource consumption (i.e., the efficiency of control processes) is lost.

The literature on the validity of the WCST is huge, mainly because the test has become the benchmark index of executive functioning. In their review, Heaton et al., (1993) demonstrate that the WCST has been shown to distinguish focal brain damage, diffuse brain damage, frontal lobe lesion, schizophrenic, seizure disorder, parkinsonian, ADHD, Tourette syndrome, autistic disorder and learning difficulty groups from well-matched normal groups. Note, however, that this suggests that the discriminant validity of the WCST (the extent to which it can distinguish one aspect of executive functioning from another) must accordingly be quite low: the cognitive and behavioural impairments associated with the disorders listed above are multiple and disparate (Ozonoff, 1995).

Also, several recent studies have suggested that when participants' abilities in the other cognitive domains are taken into account, the WCST does not reliably discriminate patients with focal brain damage from those with diffuse damage, nor does it distinguish patients with focal frontal damage from those with focal non-frontal damage (Anderson, Damasio, Jones & Tranel, 1991). Other authors have commented that some perfectly normal participants may perform very poorly on the WCST (Nelson, 1976), while some very disturbed "dysexecutive" patients can perform to a high standard (Shallice & Burgess, 1991). Heaton et al., (1993) admit, "the frontal-nonfrontal differences that have been reported do not appear to be sufficiently robust to warrant using the WCST as a frontal-lobe sign when diagnosing individual patients." Heaton et al., (1993) also note that, in their large standardisation sample, performance correlates both with age and with years of education. This is in line with Duncan's (1995) suggestion that executive functions are a property of fluid intelligence, which tends to decrease with age

3) The Task Set Approach

a) Background

It may be seen that contemporary approaches to executive function tend to emphasise high task specificity in control processes. Interpretation and modelling of the "executive function" tasks outlined above, since they are so dependent on other cognitive systems, benefits from fractionation within the task domain: into components of task-set (the relationship between the stimulus and the response) and task-set configuration (the organisation and representation of task elements) or task-set inhibition (the isolation of a task set from control over responses). For these reasons, tests involving task-set shifting and task-set inhibition have particularly attracted the attention of experimental psychologists in recent years. This advance has been paralleled by increasing interest in the development of executive functions in children, for whom acquiring control of cognition and action constitute significant developmental and social tasks. I shall outline the task set approach to control in more detail here before discussing recent developments in the understanding and assessment of executive functions in children.

Welsh and Pennington (1988) have operationally defined executive functioning as including, "the ability to maintain an appropriate problem-solving set for the attainment of a future goal." This mental set might involve any of the following: the intention to inhibit a response or defer it to an appropriate time; the planning of appropriate action sequences; or a mental representation of the task at hand, including stimulus information and the desired goal state. According to their muscular definition, executive function is relatively distinct from the other cognitive domains of language, memory, sensation, perception and even attention. In a sense it is the aspect of cognition that integrates these domains, and comes between them and action.

Monsell (1995) notes that to configure one's mental processes to perform one task rather than another is to adopt a task-set, and task-set configuration entails the organisation and linking of a chain of component processes; including, say, categorisation of sensory input, mapping the attribute's value to a response category (via a decision criterion) and execution of an appropriate motor response. For example, when approaching traffic lights in a car the driver decides whether the light is red, amber or green. If it is red, the driver recognises that this signal means "Stop". With any luck the driver applies his foot to the brake pedal. Notice that task-set configuration can occur with or without conscious "awareness": experienced drivers do not have to "think" about lights and brakes, the responses can happen automatically.

Task-set configuration is typically viewed as being endogenously (internally) driven by the executive; often in advance of a stimulus and as a preparation for action. A good example of this purposive adoption of task-set is the preparation induced by the instructions typically given to experimental participants e.g., "Start when you hear the first bell." Executive processes are also required to shift between the different tasks afforded by the same stimulus. For example, in the instruction, "Start when you hear the first bell; stop when you hear the second bell," the same bell is associated with two different responses. Executive processes are required to reconfigure mental processes and their organisation such that one particular stimulus-response mapping rather than another will be effective. For a more everyday example, consider that in a school the ringing of a bell will mean different things to pupils depending on the time of day (lunch or double maths perhaps), and whether it is a short ring or a continuous alarm.

There is also evidence that certain stimuli can in themselves activate task-sets with which they are strongly associated by factors such as frequency, recency, and salience. For example, wherever you are (school, shop, office) the sudden onset of a continuous ringing bell tells you that something is wrong and you must do something about it. The most moving examples of this exogenous (external) control are probably those of "utilisation behaviours" in some individuals with damage to the frontal lobes (e.g., Lhermitte, 1983; Shallice, Burgess, Schon & Baxter, 1989). These patients are pathologically unable to inhibit the execution of behaviours triggered by objects with which they are habitually, and normally, associated.

However, more mundane examples of this “capture of control” are common in everyday life. Reason (1990) collected many comical examples of these “action slips”, and they are recognizable to anyone who has put the sugar bowl in the fridge, tried to open their front door with the car key, or picked up the telephone when the door-bell rang. On the other hand, experimental studies of task set configuration have been rare. However, recent advances in the methodology of task-set shift and task-set inhibition experiments have substantially elucidated these aspects of control

b) Task Set Shift

Task-set shifting procedures typically require participants to shift predictably between one task and another. This paradigm was pioneered by Jersild (1927), replicated by Spector and Biederman (1976), and elaborated by Allport, Styles and Hsieh (1994). These authors required their participants to perform one of two tasks in response to each of a series of stimuli (e.g., adding versus subtracting 3 from a number, or colour- versus word-naming a Stroop stimulus). They compared performance on same-task blocks (e.g., always colour naming) and shift-task blocks (e.g., alternating between colour-naming and word-naming on each trial; ABABAB). The core finding is that, in normal adults, there are substantial costs to response time and error rates in the shift-task condition relative to the same-task condition. However, these findings are difficult to interpret since, in each case, the costs of shifting between tasks is confounded with the costs of keeping two task-sets active in the shift-task blocks (versus only one in the same-task blocks) of trials. Note that this is also true, for example, of part B of the Trail Making Test.

To resolve this issue, Rogers and Monsell (1995) developed an "alternating runs" paradigm for the investigation of task-set configuration, in which the participant must shift task-set predictably within a block of stimuli (e.g., after every two trials, AABBAABB). This paradigm allows shift-task trials (AB or BA) to be compared with same-task trials (AA or BB) in blocks where both task-sets are to remain active. Rogers and Monsell presented participants with character-pairs containing a letter or a digit and a neutral character, or both a letter and a digit e.g., G%, $6 or A2. They showed that when participants were required to shift predictably between two simple cognitive tasks, letter-identification and digit-identification, there were significant increases in response latency and error rate. These performance costs were resistant to practice and were attenuated but not abolished by long inter-trial intervals. Moreover, the shift-task costs were greater when there was "cross-talk" between the two tasks i.e., when the stimulus pair included a task-irrelevant character that was associated with a response in the other task (such as B5 or 4F).

Rogers and Monsell conclude that task-shifting comprises both a time-limited, endogenous reconfiguration component, and a stimulus-dependent, exogenous component. Task-set configuration is also vulnerable to activation of competing tasks afforded by the same stimuli. Though this attractively simple task-set shift paradigm is clearly a first attempt at elucidating the nature and mechanisms of cognitive control, it has already proved valuable at teasing apart some of the components of that control, and goes some way towards separating "task performance" from "task-set configuration". The paradigm also shows that task configuration processes are time-dependent and consume resources in ways which can be reflected in response times and error rates on simple cognitive tasks. This methodological development offers a significant opportunity to examine the executive functioning in some forms of psychopathology and neuropsychological disorder: it involves basic cognitive tasks in an uncomplicated rubric, and is therefore particularly suitable for use with younger or delayed children.

c) Task-set Inhibition

Task-set inhibition procedures are those which require the participant to withhold responses to certain stimulus configurations. The Stroop task may be framed in these terms: the participant must respond according to one stimulus attribute (ink colour) and inhibit responses according to another (lexicality). Purer inhibition tasks require the participant to respond on some trials within a test and not on others. The standard Go-NoGo reaction time paradigm in experimental psychology involves task-set inhibition, especially when the proportion of Go trials greatly exceeds NoGo trials, and strategy is weighted in favour of responding as quickly as possible. Logan (1994) has extended this approach to studying inhibition by developing the Stop-signal Paradigm, where a response is required to a primary task (e.g., letter identification) unless a secondary signal (e.g., an auditory tone) is presented.

Logan has shown that primary task response times and error rates increase significantly under conditions employing a stop-signal. For normal adults, the costs to performance of a stop-signal are impressive when signal onset is less than or equal to the mean response time for trials of the same task without signal expectation: the participant must withhold their response until the (estimated) signal onset time has elapsed. The ability of participants to adjust their response time strategically in relation to stop-signal onset has been explicitly modelled as an executive function. The executive must configure and maintain the primary task-set (e.g., "read the word aloud as quickly as possible") and, simultaneously, a second task-set (e.g., "detect the bleep") which qualifies the first (i.e., "don't read the word aloud"). For all trials the primary task response must be deferred, and on some trials it is totally inhibited.

Stop-signal onset is usually calibrated individually for each participant. This is to ensure (as if in a single-channel "horse race" between primary task processing and the signal onset) that signal presentation and the participant's readiness to respond typically coincide (Logan, 1994; Pashler, 1993). Logan's stop-signal paradigm has yet to be employed in studies on patients with frontal lesions, but as suggested in the discussion of utilisation behaviours, it is a common clinical observation that frontal lobe patients have difficulty inhibiting or deferring well-learned or primed behaviours (Lhermitte, 1983). Inhibiting and deferring certain behaviours is also a recognizable quality of young children, and a persistent difficulty in children with profound developmental psychopathology. As we shall see, researchers in this field have shown increasing interest in the relationship between the development of executive functioning, and measures of task-set shift and inhibition.

4) Development of Executive Function and Dysfunction

a) Background

Relatively little is known about the normal development and nature of executive functions in children. In his review Stuss (1992) suggested that the frontal systems (and associated executive functions) are among the last to mature in normal development. Before birth, cortical development proceeds from the primary to the secondary sensory and motor areas, then to the association regions, with the pre-frontal cortices developing last. Formation of the cortical gyri and laminar structure of the frontal regions is nevertheless almost complete at birth. Morphologically, the frontal regions are the last to mature: structural pyramidilization, metabolic activity, electrophysiological patterning and cellular myelination in the frontal lobes are not comparable to the anterior regions until around puberty. Stuss makes the (not astonishing) suggestion that development of the frontal systems depends on the interaction between biological maturity and empirical experience. Recently, the psychological aspects of control functions in children have engaged more attention because some of the profound developmental psychopathologies have attracted investigation and explanation in terms of executive dysfunction.

b) Executive Dysfunction in Children

Using the prescription tests (outlined above) executive function deficits have been identified in groups of children diagnosed as having Autistic Disorder, (e.g., Ozonoff & McEvoy, 1994), Attention Deficit Hyperactivity Disorder (ADHD; e.g., Barkley, Grodzinsky & DuPaul, 1992), Conduct Disorder (Moffitt, 1993), Gilles de la Tourette Syndrome (e.g., Baron-Cohen, Cross, Crowson & Robertson, 1994), early treated Phenylketonuria (e.g., Welsh, Pennington, Ozonoff & Rouse, 1990), Sydenham's Chorea (Casey, Vauss, Chused & Swedo, 1994) and Turner's Syndrome (Temple, Carney & Mullarkey, 1996). In their review and meta-analysis of nearly 70 papers in this area, Pennington and Ozonoff (1996) show that studies of executive dysfunction employing the WCST, Stroop latency, Go-NoGo tasks, the Tower of London test, Trail Making part B, and Working Memory tasks have consistently discriminated groups of children with Autistic Disorder or ADHD from normal and learning difficulty samples. Pennington and Ozonoff note that the literature on Tourette Syndrome is presently inconsistent with regards to executive function deficits, while the evidence from well designed studies of Conduct Disorder (where co-morbidity with ADHD is controlled) is overwhelmingly negative. To date, too few studies have been conducted on samples of children with the other pervasive disorders diagnosable in childhood for patterns have been detected.

The evidence from biological investigations of brain function is compatible so far with the findings from clinical neuropsychology. In their review, Pennington and Ozonoff conclude that there is strong evidence for decreased blood flow to the frontal lobes in ADHD children, whereas hyper-frontality of blood flow is characteristic of the normal brain. Other evidence suggests that the basal ganglia are the loci of decreased blood flow in ADHD, while flow is increased to the motor and sensory cortices. The construct of a frontal lobe dysfunction in ADHD has also been based on the observation that frontal lesions, in animals and human patients, tend to result in hyperactivity, distractibility, impulsivity, or deficits in motor control, which are the core behavioural symptoms of ADHD. Similarly, good evidence exists for depressed cerebral blood flow to the right-, left- and mid-frontal lobes in adults, adolescents, and younger children with autistic disorder. Autistic children also show less frontal power in EEG examination than matched controls. The biological evidence for frontal system dysfunction in Tourette syndrome is mainly metabolic. Decreased glucose metabolism has been found in the middle, orbital and inferior regions of the frontal lobes in adults with the syndrome, but reduced cerebral blood flow to frontal cortex has also been reported, and as in ADHD the basal ganglia is a putative focus of this deficiency.

However, *that* executive function and frontal-lobe deficits have been identified in these groups has perhaps proved more interesting to researchers than *why* this might be so. Most of the studies summarized above are innocent of any detailed model of the relationship between executive dysfunction and core or peripheral psychopathology. Barkley, Grodzinsky and DuPaul (1992) proposed an inhibition deficit model of ADHD: a cognitive model linking executive deficits in inhibition directly to the behavioural disturbances of hyperactivity, distractibility, and impulsivity. The core disturbances in Tourette syndrome include deficits in sensory, motor and visuo-spatial functions, but Gedye (1991) has suggested that the movements and utterances typical of this disorder may reflect failure of a motor inhibition system, mediated by the "central executive". Unfortunately, is not clear is how these two models may be adequately distinguished.

Some of the core disturbances in autistic disorder and Asperger's syndrome map directly onto prototypical executive deficits, especially perseverative and stereotyped behaviours, resistance to novelty and change, limited range of interests, and impaired abstract thought. However, other sophisticated models of the core psychopathology in these disorders already exist e.g., impaired mental state processing (Theory of Mind; Baron-Cohen, Leslie & Frith, 1985), or impaired affective and interpersonal relatedness (Hobson, 1989). The main issue for an executive function theorist therefore would be identify the underpinnings of elements of these models in an executive deficit terms. Accordingly, Ozonoff (1995) suggests that impairment in the deployment of "if-then" rules would impair performance on both executive function and Theory of Mind tasks. Other authors have suggested that a failure to disengage from contact with the immediate environment would explain the pretend play and deception task impairments of autistic disorder children. Further investigation of the specific associations between distinct executive functions and particular behavioural disturbances are urgently needed if powerful, inclusive models of the developmental psychopathologies are to be generated.

c) Assessing the Executive Functions in Children

The established neuropsychological tests of executive function are not easy to use with children, and accordingly their results are difficult to interpret or extend to domains of psychopathology. Those tests which depend on levels of literacy (Verbal Fluency), reading ability (classic Stroop) or general knowledge (Cognitive Estimates Test) produce artificially poor and highly variable performance even in normally developing children. It is not clear whether the abilities they measure (reasoning about qualitative variables, reading colour words, producing words to an initial letter) better reflect linguistic or associative abilities than the executive functions, and are therefore unduly influenced by (wide) variations in children's literacy and educational experience. These tasks cannot be used with confidence on pre-literate children, and therefore have a technical participant floor at around age seven.

As already indicated, the Wisconsin Card Sort Task can be very difficult to administer to children because of its complexity and rather aversive rubric. For all standard neuropsychological tests of executive function (including the WCST) adequate norms do not exist for the youngest age groups, i.e., below six or seven years of age. Recently, several authors have conducted preliminary research using novel Go-NoGo, Stop-signal, and Task Shift paradigms in studies of the development of executive functions in children. Factors affecting the control or inhibition of behaviour have been investigated with these procedures, both in normally developing children and those with profound developmental psychopathologies, with some success.

Hughes, Russell and Robbins (1994) employed a computerized intradimensional-extradimensional set-shifting task to compare executive functioning in an Autistic Disorder sample with matched normal and mild learning disability (MLD) groups. Participants were presented with two boxes containing one of four white lines superimposed upon one of four pink shapes. As in the WCST the participant must deduce from the pattern of reinforced responses which of the designs is “correct”. At first it is always one of the pink shapes, and the white lines are to be ignored. The criterion for rule change is eight correct responses in a series of 50. On intradimensional (ID) shift trials, the rule changes so that a different pink shape becomes the correct one i.e., ID shift entails transfer to a new exemplar of the same form. On extradimensional (ED) shift trials, the rule changes so that the correct design is now given by one of the white lines i.e., ED shift entails transfer to an exemplar of a different design element.

On ID shift, the Autism group showed significantly more "errors to criterion" than the MLD and normal groups, suggesting they had difficulty in transferring learning to new exemplars of the same form. On the ED shift, the Autism group were relatively impaired on all measures of performance (errors to criterion, trials to criterion, perseveration of response). The pattern of results suggests that children with autistic disorder have impaired executive functions involving set shifts *between* dimensions more than *within*. However, like the WCST the ID-ED paradigm depends on a rule-deduction procedure (task-set learning) in addition to set maintenance and shift functions, so impairments on this tests might be as well related to task demands on learning and memory as to the executive functions. Hughes et al., did not detect significant performance trends associated with chronological or criterion-approach development on their task.

In contrast, Diamond and colleagues have shown that clear developmental trends are detectable, in normally developing children, on tests of cognitive and motor inhibition. Gerstadt, Hong and Diamond (1994), examined the performance of normal children, between 3 ½ and 7 years, using a "response-opponence" inhibition paradigm. In this case the task was to say "day" to a picture of the moon and "night" to a picture of the sun: the participant must inhibit a pre-potent response in favour of its conventional opposite. Children under 5 years performed poorly on this task, making more errors and taking longer to respond as trials continued. Children at all ages performed better in the earlier trials than in the later ones, suggesting that the efficiency of inhibitory functions is time-dependent. Response latencies and error rates improved with age such that children of 6 and 7 years performed at the same high level throughout the procedure.

Similarly, Diamond and Taylor (1996) examined these processes of inhibition in the performance of normal children, between 3 ½ and 7 years, on Luria's tapping test. In this test the participant must tap once if the examiner has tapped twice, and tap twice if the examiner has tapped just once. The participant must hold these two rules in mind (e.g., "tap twice when I tap once") and exercise inhibitory control over pre-potent behaviour (i.e., resist the urge to copy the examiner.) Children of 3 and 4 years performed poorly on this task, making more errors and taking longer to respond than children of 5 years and older. Again, all children performed better in the earlier trials than in the later ones. Response latency and error rates decreased with age such that children of 6 and 7 years performed at a near ceiling level throughout testing. The pattern of result in these studies suggests that some executive functions develop rapidly between 3 and 5 years of age, and are fast, stable, and efficient by 6 or 7 years.

Ozonoff and colleagues have compared the performance of children with autistic disorder, and Tourette syndrome, with normally developing groups on some sophisticated cognitive tasks. Ozonoff, Strayer, McMahon and Filloux (1994) examined autistic disorder and Tourette syndrome groups on a Go-NoGo test involving three conditions. Participants were presented with a square or circle on the computer screen. In the first condition (neutral inhibition) one of the figures was the Go stimulus, and one the NoGo stimulus: participants were required to press the space bar on a key board whenever they saw the Go stimulus. In condition 2 (pre-potent inhibition), the preceding Go stimulus became the NoGo stimulus and vice versa. In condition 3 (flexible shift of inhibition) whether the circle or square was the Go stimulus varied across blocks of trials. Ozonoff et al., found that the performance of children with Tourette syndrome did not differ from matched controls on any of the inhibition conditions. Autistic children, however, showed impaired performance on the pre-potent condition and markedly impaired flexible shift of inhibition.

Ozonoff and Strayer (1997) examined the performance of children with autistic disorder on Stop signal and Negative Priming tasks. In the stop-signal task, participants were required to categorize words as animal or non-animal names, unless an auditory tone indicated that they should suppress their response. In developing the Negative Priming paradigm, Tipper (1985) found that when a target stimulus on trial N had been the ignored stimulus on trial N-1, participant response times are slowed and error rates increase. (This "negative priming" effect occurs even when participants subsequently show poor memory for the ignored stimuli.) Tipper suggested that the representations of ignored stimuli are actively inhibited: inhibitory processes isolate the ignored distracter codes from the control of responses, and it is more difficult to respond according to these codes on subsequent trials.

In their negative priming task, Ozonoff and Strayer presented participants with a five letter string (e.g., FTFTF) and required them to judge whether the second and third letters were the same or different. On negative priming trials, the target letters had been the distracter letters on the previous trial. Ozonoff and Strayer found that autistic disorder children were not impaired, relative to matched controls, on either the Stop-signal or Negative Priming conditions. They conclude that modularity within the executive functions is paralleled by different patterns of dysfunction in the disparate developmental psychopathologies.

To summarize, the available evidence suggests that some sophisticated cognitive tests are efficient in detecting patterns of both preserved and impaired abilities within the domain of executive functions. Also, there is fractionation within the domain of psychopathology: disparate executive functions appear to be affected by different disorders. Notice that most of the studies outlined above compared groups of psychological disorder children with groups of normal controls without considering the effect of developmental trends within the groups. There is good reason to suspect that (as Diamond and colleagues have shown in normally developing children) executive functions may develop over time even in children with profound psychopathology. One way to examine this issue would be to adopt a "battery approach" to assessing abilities: a range of executive functions could be investigated (relative to each other) within a participant. At the individual level, it is important to pay attention to developmental trends, not only in success on a task but also in the speed at which it can be done.

6) A Computer-based Assessment Battery

The finding that methods derived from cognitive psychology might be useful indices of executive dysfunction, in certain psychological disorders, is important for several reasons. At a more theoretical level, it seems likely that executive dysfunction may be a factor precipitating or maintaining some profound developmental psychopathologies. Thus, if different disorders have specific effects at distinct loci in the executive system then this may help in the classification or modelling of disorders, and perhaps in the generation of novel interventions. At a lower level, these methods might provide more objective means of screening for disorder, achieving a differential diagnosis, or monitoring the progress of an intervention.

Several tasks from the cognitive literature appear good candidates for inclusion in an executive function assessment battery. If task-set inhibition is a function of interest, then the response opponence (day-night) and stop-signal tasks, having been shown to capture normative developmental trends and specific executive deficits, should be included. If task-set shifting is of interest, then a procedure resembling the alternating runs paradigm (Rogers & Monsell, 1995) would be useful. Executive functions can be further stretched in this paradigm by including “cross-talk” stimuli, which afford a response to more than one task-set. Furthermore, an investigation of the effect of exogenous (external, stimulus-cued) control is lacking in the literature. It would be useful to compare the effect of normal development and profound psychopathology on performance when set-shift is endogenously “prepared” versus set-shift that is stimulus- or cue-dependent.

While the established neuropsychological tests of executive functioning are traditional "paper and pencil" tasks, the novel paradigms tend to require computer-based administration and scoring. Computer-based measurement of responses to single-stimuli are also the most useful way to investigate the properties of assessment batteries. Technically, computer-based measurement of individual latencies has a number of advantages over measurement of, say, the time taken to complete a whole block or test. Greater temporal precision is achieved and errors can be excluded from the latency analysis for separate investigation. Also, trial by trial presentation prevents participants from attending to the previous item, or previewing the next item, while completing processing of the current one. Finally, computer-based presentation of materials has the distinct advantage of partialling out the interpersonal effect of the examiner on the participant's performance: this might be crucial where the main or adjunctive aspects of a disorder are difficulties in social interaction (Ozonoff, 1995).

The present study is a report on the development of a computer-based assessment of children's executive functioning, and provides preliminary data on its use. The immediate aim is to investigate normal executive functioning, using a cross-sectional design, across a brief developmental period (7 to 11 years) and to suggest which abilities are, or become, faster or more accurate over this period. The functions examined are: those involved in simple task-set configuration, task-set configuration under a condition of stimulus "cross-talk", task-set configuration under a condition of stimulus-cued control, and task-set inhibition. The assessment battery outlined below involves a suite of task-set configuration and shift conditions (based on the alternating runs paradigm; Rogers & Monsell, 1995) and a pair of inhibition tasks.

The task-shift series involves just two simple tasks: naming the shape of a figure or naming the colour of a colour-patch. Colour-naming and shape-naming were selected, in preference to reading or counting tasks (or tasks requiring a manual response), because they are fundamental educational sets which are nevertheless relatively independent of literacy and numeracy skills (Bornstein, 1985; Davidoff, 1991). Unlike in the WCST or the ID-ED paradigm, in the present design participants do not have to deduce the rules from the pattern of reinforced responses. Rather, the rules are given verbally by the experimenter, participant comprehension is checked, and the tasks are practised to ensure competence.

Initially, colour-naming and shape-naming are practised separately to ensure that task performance is consolidated. Next, task-set configuration is examined under a condition of shifting between the two tasks in an alternating runs design: the participant must colour-name a colour patch for two trials, then shape-name a figure for two trials and so on. Executive functions are further taxed in the next condition by including stimulus cross-talk. The stimulus is a coloured shape, which affords a response to either task-set: it could be colour-named or shape-named, but task order is given by the alternating runs design. Task-set configuration is then examined under a condition where the shift between task-sets is unpredictable. The stimulus itself cues the appropriate response (exogenous control): a colour-patch cues colour-naming, a figure cues shape-naming. These task-shift tests produce measures of the costs to response time and error rates of shifting between task-sets.

The first of two inhibition tests is a simple response-opponence condition based on Gerstadt et al's (1994) day-night task. The participant must inhibit a well-learned response in favour of its conventional opposite. However, the Gerstadt et al., design entailed children making additional associative connections of "sun" with "day" and "moon" with "night". In the present version, the participant has only to say "sun" to a picture of the moon, and "moon" to a picture of the sun. This approach is intended to make the task easier to comprehend, but involves more direct (and therefore stronger) interference between the stimulus dimension and required response. Since the participant is required to respond on every trial, it is assumed that response times and error rates will primarily reflect processes of cognitive inhibition.

Finally, a Stop-signal task is given, where the participant must inhibit responding on certain trials. The participant is required to name the colour of a colour-patch as quickly but accurately as they can unless a stop-signal (a black cross within the patch) is presented: the black cross means that the participant is to remain silent. Unlike in the standard Stop-signal design, which uses an auditory tone to suppress responding, the primary stimulus and stop-signal are integrated in the present design. This allows the participant to sustain attention to one region in one sensory modality: the participant does not have to search the visual array, or divide attention with the auditory modality, in order to detect the inhibitory cue. Stop-signal onset is individually calibrated for each participant, and it is assumed that response times and error rates will therefore reflect processes of motor inhibition.

This fixed presentation order (colour naming, shape naming, simple shift, cross-talk shift, cued shift, sun-moon, stop-signal) was chosen for this preliminary study for several reasons. First, if a standardized assessment battery is to be developed, then this will necessarily be presented in the same fixed order to all participants. Second, for the set-shifting blocks, performance is expected to depend on well-practised and efficient colour and shape naming performance and these are accordingly given first. Similarly, practice on task "location" (upper boxes for colour naming, lower boxes for shape naming) is assumed to be necessary for assessment of set shifting under conditions of stimulus cross-talk (when stimulus location is the main cue to the appropriate task set) so that the simple shift condition is accordingly given before the cross-talk block. Furthermore, practice on several blocks where all trials require a response was assumed to be a useful way of priming response tendencies before challenging them in the stop-signal task, and this was accordingly given last.

Method

1) Participants

Forty children from seven to eleven years of age (five boys and five girls from school years 3, 4, 5 and 6) were recruited from a mixed junior school. Class teachers were asked to select a sample of five boys and five girls from their group who were native English speakers, of average ability, without known or suspected medical, social, familial or behavioural difficulties. An information letter and consent form was sent, via the child, to the parents/guardians, two weeks before testing took place. All children who were invited to participate agreed and were tested. Ethical approval for the study was given by the Joint University College London and University College London Hospitals Committee on the Ethics of Human Research. Copies of the approval notice, information sheets and consent forms, for parents and children, are provided as Appendices.

2) Design and Procedure

After giving informed consent, participants were tested individually in a quiet room. First, estimates of each child's verbal and performance abilities were obtained using the vocabulary and block design sub-tests of the Wechsler Intelligence Scale for

Children Third Edition UK (WISC-III^{UK}, 1992). Two WISC sub-tests were chosen since they are relatively brief, and allow for a broad range of ability to be measured. WISC vocabulary correlates best among the other verbal sub-tests with both overall Verbal functioning (r = .82), and the scale's Verbal Comprehension factor (r = .85) across all age groups. WISC block design correlates best among the other performance sub-tests with both overall Performance functioning (r = .76) and the scale's Perceptual Organisation factor (r = .78) across all age groups (Wechsler, 1992). Also, parallel versions of these two sub-tests are used in the Wechsler Pre-school and Primary Scales of Intelligence (WPPSI-R, 1990) and the Wechsler Adult Intelligence Scales (WAIS-R, 1981) which would facilitate future comparison with older or younger population norms.

Next, the participant sat with their eyes approximately 50 cm from the screen of a Macintosh LC II computer. The participants wore earphones (to attenuate external noise) and a boom microphone connected to a voice key, interfaced to the computer. Naming latencies were measured to the nearest millisecond. The experimenter sat behind or next to the participant, supervising the presentation of blocks, monitoring the operation of the voice key, and recording incorrect, disfluent, and undetected responses on a score sheet. Training and practice on task requirements, using printed paper versions of the up-coming block, were given immediately prior to each block.

The assessment battery comprised seven blocks of 44 trials. The first four trials in a given block were counted as "run-in" or practice trials and were not included in the analysis. The blocks were presented in the same order to each participant, and there was a short break between each block and the next. In each break, participants were praised for the very few errors they had made and for the speed of their performance (regardless of actual performance success) and were encouraged to work as quickly and accurately as they could.

For the first four blocks a framework of four boxes was displayed on the computer screen, the side of each box being 5 cm. A single 5 cm box was displayed for blocks 5, 6 and 7. The boxes were printed in white against a black background. On a given trial a colour-patch or figure was presented individually in one of the boxes. On colour-patch trials, the whole box was filled in one of four colours (red, yellow, green or blue). On figure trials, a shape (circle, triangle, square or star) or a picture (sun or moon) was presented horizontally- and vertically-centred in the square. Figures were sized to fill an area of 3 cm^2. The stimulus remained visible until the participant responded or 3000 msec has elapsed. The next stimulus was presented 1000 msec after the detection of a response to the previous one.

In block 1, the colour-naming task, the task was to name the colour in which the box was filled. The colour appeared in one of the two upper boxes; first the left, then the right, then the left and so on. Participants were told that a colour would fill one of the two upper boxes, and that they should name the colour as quickly but as accurately as possible. In block 2, the shape-naming task, the task was to name the shape of a figure: four shapes (circle, triangle, square and star) were line-drawn, unfilled, in white against a black background. The shapes to be named appeared in one of the two lower boxes; first the right, then the left, then the right and so on. Participants were told that a figure will appear in one of the two lower boxes, and that they should name its shape as quickly but as accurately as possible. Blocks 1 and 2 are colour- and shape-naming practice blocks, which provide baseline measures of the preparation and maintenance of well-practised task-sets.

Block 3 was a simple shift-set condition, where the task-set (colour- versus shape-naming) was shifted every two trials giving response latency and error rate measures of the costs of shifting between two tasks. The task was to name the colour of the box for two trials, then the shape of the figure for two trials, and so on in an alternating runs design. Stimuli appear in all four boxes in turn, starting with the top left and going clockwise around (top left, top right, bottom right, bottom left). Task order was therefore given by stimulus position on the visual display: upper boxes for colour-naming, lower boxes for shape-naming. Participants were told that when a colour appeared in one of the upper boxes they should name the colour, and when a figure appeared in one of the lower boxes they should name its shape, as quickly but as accurately as possible.

Block 4 was a cross-talk shift-set condition: again the task was to shift between colour- and shape-naming in an alternating runs paradigm, but the four shapes (circle, triangle, square and star) now appeared line-drawn in the four colours, in all four boxes in turn (starting with the top left and going clockwise around). Task order was still given by stimulus position, but not stimulus attributes. Thus there was "cross-talk" between the two tasks since each stimulus now afforded a response according to either task-set: it was possible to colour-name or shape-name any stimulus. Participants were told that coloured figures would be appearing in all four boxes: when a figure appeared in one of the upper boxes they should name its colour, but when a figure appeared in one of the lower boxes they should name its shape, as quickly but as accurately as possible. Note that the alternating runs design in blocks 3 and 4 provided 22 "shift task" and 22 "same task" trials.

Block 5 was a cued shift-set condition: the task was to name the colour of any colour-filled box that appeared, and to name the shape of any figure that appeared (the shapes always appeared printed in white). The framework of four boxes was replaced by a single box (5 cm) in the centre of the screen. Task order was no longer given by alternating runs, but by an approximate binomial distribution of 22 "shift task" and 22 "same task" trials. The appropriate task was not cued by position but by the colour or shape of the stimulus: figures for shape-naming, filled boxes for colour-naming. Participants were told when a colour appeared they should name the colour, and when a figure appeared they should name its shape, as quickly but as accurately as possible.

Block 6 was a response-opponence condition, where the stimulus "primes" a well-learned response which must be inhibited in favour of its conventional opposite. A single box (5 cm) remained in the centre of the screen. A picture of the sun or the moon appeared on each trial. The sun picture (a filled circle with rays printed in yellow) and moon picture (a filled crescent printed in pale blue-grey) were each presented 22 times in an order given by an approximate binomial distribution. Participants were told to say "sun" when a picture of the moon appeared, and to say "moon" when a picture of the sun appeared in the box, as quickly but as accurately as possible.

Block 7 was a stop-signal condition, where the participant was required to withhold their response to a primary task if a secondary signal is detected. The task was to name the colour of a box unless a cross appeared shortly after the box had filled with colour. A single box remained in the centre of the screen. First, a colour filled the box on each trial, then on 50 per cent of trials a large black cross (the stop-signal) appeared in the centre of the colour patch. Stop-signal onset time was calculated individually for each child: colour-to-signal asynchrony was set as the mean colour-naming time achieved in block 1 minus 150 ms. Participants were told to name the colour of the box as quickly but as accurately as possible, unless a cross appeared in which case they were to remain silent.

At the end of block 7, participants were thanked for their time and, by way of a debrief, were invited to comment on which blocks they had found easiest and hardest. The mean age in months, and age-scaled vocabulary and block design scores for all participants are summarized in Table 2.1. It may be seen that the sexes were well matched on these values within the school years, and t-tests confirmed that the boys did not differ from the girls in terms of age, t (38) = 0.49, vocabulary scores, t (38) = 0.83, or block design scores, t (38) = 0.25. The overall mean vocabulary score appears to exceed the mean block design score: a paired samples t-test procedure confirmed that this difference was significant, t (39) = 3.92, p<.001, but the scores were reliably correlated, r = .64, p<.01. Accordingly, the vocabulary and block design scores were averaged for each participant to produce a general index of "ability", and these ability scores are also summarized in Table 2.1.

Notice, however, that the youngest children (year 3) achieve lower scores on average for the vocabulary and block design sub-tests than the older groups. Kruskal-Wallis one-way ANOVA procedures revealed that, across the school years, mean vocabulary scores were significantly different, H (3) = 10.233, p = .017, but mean block design scores were not, H (3) = 5.328. As these differences should not reflect properties of the WISC age-scaling formulae, then a selection bias must be assumed to be in operation over the school years. Unfortunately, this selection bias produced a significant correlation between age and ability (r = .38, p<.05) which must be taken into account if the two variates interact in statistical analyses.

Table 2.1. Participants' Age in Months, Vocabulary Scores, Block Design Scores, and Ability by Sex and School Year.

		Age in Months		**Vocabulary Scale Score**		**Block Design Scale Score**		**Ability**	
	School Year	*Mean*	*(SD)*	*Mean*	*(SD)*	*Mean*	*(SD)*	*Mean*	*(SD)*
Girls	3	96	(5)	8	(3)	7	(3)	8	(2)
	4	107	(3)	13	(3)	9	(3)	11	(3)
	5	121	(4)	12	(4)	9	(2)	10	(3)
	6	131	(5)	12	(3)	12	(3)	12	(3)
Boys	3	97	(3)	9	(3)	8	(3)	8	(3)
	4	111	(2)	11	(2)	12	(2)	12	(2)
	5	124	(2)	11	(2)	10	(2)	11	(2)
	6	133	(4)	12	(3)	8	(3)	10	(3)
Both	3	96	(4)	8	(3)	8	(3)	8	(2)
	4	108	(3)	13	(3)	10	(3)	11	(2)
	5	122	(4)	12	(3)	9	(2)	11	(2)
	6	132	(4)	12	(3)	10	(3)	11	(3)
	Overall	114	(15)	11	(3)	9	(3)	10	(3)

Results

Timing failures, where a spoken response was not detected, the voice key was prematurely tripped by an irrelevant noise, or the trial was aborted for some reason, accounted for 2.35 per cent of trials. For blocks 1 to 6, a "response error" was recorded if the participant named the stimulus incorrectly, gave a disfluent response or made no response. In addition, for the cross-talk shift condition (block 5) a "task error" was recorded if the participant named the shape when they were expected to name its colour, or vice versa. For the stop-signal condition (block 7) errors were recorded as either "failures" if the participant failed to suppress responding on stop-signal trials, or "misses" if the participant failed to respond though no stop-signal was presented.

1) Response Latency (RT) Analysis

Timing failures and participant errors were excluded from the latency analysis. The means and standard deviations for correct RTs, in each block and for a baseline average RT of colour-naming and shape-naming, are given in Table 3.1. These represent the "difficulty" of each block in terms of processing time per trial. Several broad patterns in the data are obvious to inspection. First, the basic cognitive tasks, colour-naming and shape-naming, are not equivalent: shape-naming takes an average of 173 msecs longer than colour-naming. A matched samples t-test performed for all participants confirmed that shape-naming latencies were significantly longer than colour-naming latencies, t (39) = 10.197 , p < .001. This difference must be taken into account when investigating the effect of shifting between the two tasks.

Second, response latencies for the younger groups appear to be longer than for the older groups (with the exception for year 4 in the stop-signal task). The pattern of RTs in the sun-moon task showed a steep improvement over years 1 to 2, and only a small change thereafter. Though the participant numbers in this study are rather small for regression analysis, these effects will be examined alongside any detectable influence of ability. Third, relative to a baseline average response time for colour-naming and shape-naming (mean = 709) there appear to be moderate costs to shifting predictably between the two basic tasks (block 3; mean = 753), large costs when the shift is cued by the stimulus itself (block 5; mean = 834) and substantial costs when there is stimulus cross-talk (block 4 mean = 1019).

Table 3.1. Mean Response Latencies (msecs) and Standard Deviations by Block and School Year.

	Colour-naming (block 1)		**Shape-naming (block 2)**		**Sun-Moon (block 6)**		**Stop-signal (block 7)**	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	743	(186)	934	(216)	1052	(326)	712	(216)
4	599	(80)	833	(194)	762	(119)	527	(210)
5	594	(205)	725	(233)	707	(177)	492	(153)
6	521	(124)	648	(151)	700	(243)	603	(247)
Overall	623	(174)	796	(223)	826	(280)	598	(221)

	Baseline (blocks 1 &2)		**Simple Shift (block 3)**		**Cross-Talk Shift (block 4)**		**Cued Shift (block 5)**	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	839	(196)	901	(233)	1266	(330)	959	(277)
4	716	(131)	738	(132)	978	(207)	841	(186)
5	660	(217)	671	(188)	908	(245)	778	(236)
6	584	(135)	651	(179)	840	(247)	720	(197)
Overall	709	(194)	753	(211)	1019	(311)	834	(240)

a) Inhibition Blocks

The RTs for the sun-moon task were subjected to a General Linear Model procedure (GLM; see statistical appendix) with the between subjects factor sex, and the co-variates age (in months) and ability. The GLM yielded no main effect of sex, but significant main effects of age $F(1,33) = 6.07$, $p<.05$, and ability $F(1,33) = 6.03$, $p<.05$. There were no significant interactions. The RTs for the stop-signal task were subjected to a GLM procedure with the between subjects factor sex, and the co-variates age and ability. This GLM yielded no main effects or significant interactions.

b) Set-shifting Blocks

Response latencies for the set-shifting blocks were divided into same-set trials, where the task on trial N was the same as on trial N-1, and shift-set trials, where the task on trial N was different to that on trial N-1. Same-set and shift-set trials were further distinguished according to whether the task on the trial was to name the colour or the shape of a stimulus. The mean correct response latencies for each trial type are summarized for the simple shift condition (block 3) in Table 3.2, for the cross-talk shift condition (block 4) in Table 3.3, and for the cued shift condition (block 5) in Table 3.4. Clear patterns emerge from visual inspection of the data. First, as observed earlier, with one or two exceptions the RTs for younger children appear longer than for older groups. Second, within the blocks, shifting to a different task incurs a response latency cost, so that shift-set RTs tend to be longer than same-set RTs. This cost appears largest in the cross-talk shift condition, but rather small in the cued shift condition, and varies in relative size between the colour-naming and shape-naming tasks.

Table 3.2. Mean Response Latencies (msecs) and Standard Deviations for the Simple Shift Condition by Same-Set or Shift-Set Trials and School Year.

	Same-Set Trials					
	Both Tasks		*Same Colour*		*Same Shape*	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	833	(210)	725	(159)	943	(294)
4	699	(155)	615	(93)	784	(219)
5	626	(188)	561	(172)	695	(212)
6	618	(155)	586	(169)	651	(165)
Overall	706	(196)	632	(160)	782	(253)

	Shift-Set Trials					
	Both Shifts		*Shift→Colour*		*Shift→Shape*	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	970	(273)	929	(300)	1009	(288)
4	776	(116)	714	(126)	840	(122)
5	715	(198)	645	(198)	788	(209)
6	685	(210)	634	(200)	737	(243)
Overall	801	(238)	747	(250)	856	(248)

Table 3.3. Mean Response Latencies (msecs) and Standard Deviations for the Cross-Talk Shift Condition by Same-Set or Shift-Set Trials and School Year.

	Same-Set Trials					
	Both Tasks		*Same Colour*		*Same Shape*	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	1173	(377)	1031	(380)	1317	(476)
4	884	(226)	810	(226)	964	(289)
5	864	(225)	841	(245)	890	(245)
6	791	(278)	755	(261)	828	(306)
Overall	947	(324)	872	(304)	1025	(399)

	Shift-Set Trials					
	Both Shifts		*Shift→Colour*		*Shift→Shape*	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	1364	(293)	1355	(295)	1367	(376)
4	1083	(196)	1097	(146)	1072	(274)
5	954	(275)	977	(329)	931	(232)
6	896	(233)	849	(248)	940	(231)
Overall	1097	(312)	1090	(323)	1102	(339)

Table 3.4. Mean Response Latencies (msecs) and Standard Deviations for the Cued Shift Condition by Same-Set or Shift-Set Trials and School Year.

Same-Set Trials

	Both Tasks		*Same Colour*		*Same Shape*	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	966	(308)	775	(194)	1129	(394)
4	803	(218)	676	(96)	930	(358)
5	764	(249)	686	(234)	836	(260)
6	694	(169)	635	(178)	758	(204)
Overall	819	(259)	699	(182)	930	(341)

Shift-Set Trials

	Both Shifts		*Shift→Colour*		*Shift→Shape*	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	960	(288)	859	(241)	1051	(321)
4	865	(181)	760	(156)	969	(239)
5	785	(231)	725	(261)	843	(211)
6	730	(213)	709	(179)	759	(276)
Overall	844	(245)	771	(213)	916	(286)

naming trials much less markedly so $F(1,39) = 6.21$, $p<.05$, though both effects were reliable. There was no significant three way interaction $F(2,78) = 1.42$.

c) The RT Costs of Set-Shift

In order to investigate the relationships between these set-shift costs and developmental level or ability it is important to consider whether the effects of task-set configuration are confounded with those of task performance in the RT data. For example, from Table 3.1 it is clear that the performance of children on simple cognitive tasks (colour- or shape-naming) improves (RTs become shorter) with age. It is a nice question, therefore, whether the observed costs to shifting between two tasks vary with age as a function of executive development or as a function of increased performance speed. Thus, shift costs can be measured either in terms of the *absolute* value RT cost (i.e., the difference between shift-set RTs and same-set RTs) or in terms of a *relative* increase in processing (i.e., shift-set performance as a proportion of same-set performance). To address this question, in this study the costs of shifting between tasks were calculated for each participant, condition and task in two ways. Absolute costs were calculated using the formula,

$$\textit{absolute RT cost} \quad \textit{(mean shift-set RT} - \textit{mean same-set RT)}$$

Relative costs were calculated using the formula,

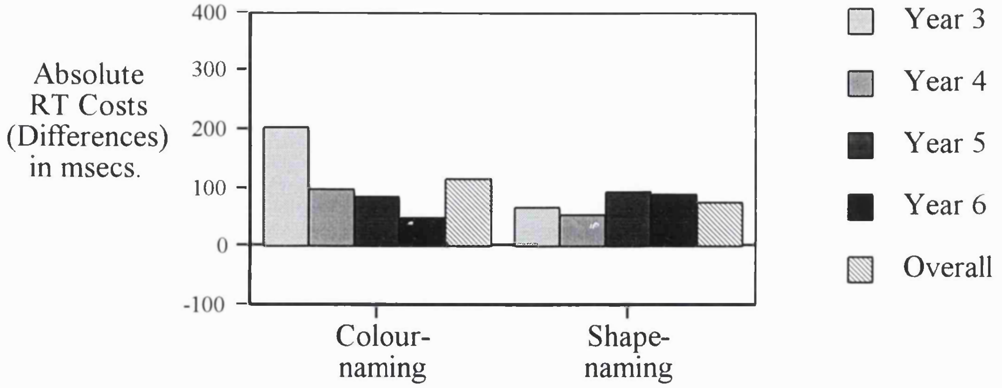
$$\textit{relative RT cost} \quad \frac{\textit{(mean shift-set RT} - \textit{mean same-set RT)}}{\textit{mean same-set RT.}}$$

The absolute RT costs are plotted for each year group and condition in Figure 3.1, while relative RT costs are plotted in Figure 3.2

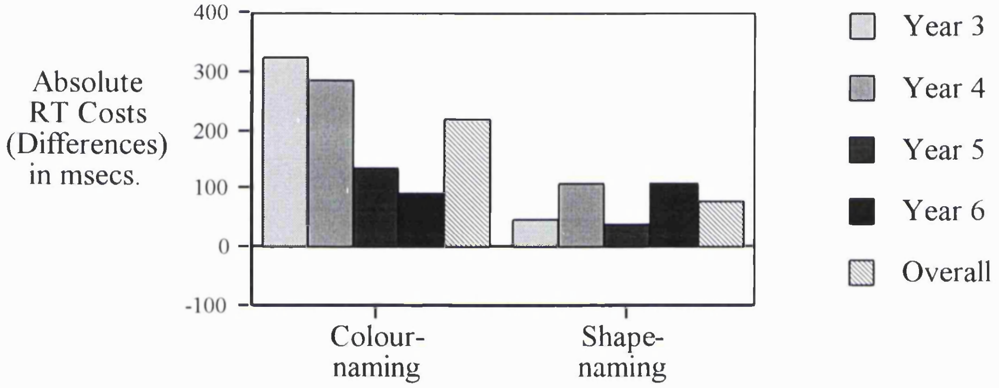
To investigate whether these differences between conditions, set-shift and naming tasks were reliable, the mean RTs for each participant were subjected to a repeated measures ANOVA with the factors condition (simple, cross-talk and cued shift), set-shift (same-set, shift-set) and task (colour-naming, shape-naming). This yielded main effects for condition F(2,78) = 76.63, p<.001, set-shift F(1,39) = 55.13, p<.001, and for task F(1,39) = 45.40, p<.001. Planned comparisons (orthogonal linear contrasts with the error term pooled across all contrasts) revealed that RTs for the cross-talk condition were reliably larger than for the simple shift condition F(1,78) = 143.82, p<.001, and for the cued shift condition F(1,78) = 74.97, p<.001. Similarly, the RTs for the cued shift condition were reliably larger than for the simple shift condition F(1,78) = 11.12, p<.01.

However, these main effects were qualified by interactions. There was a significant interaction between condition and set-shift F(2,78) = 9.11, p<.01. Simple effects tests revealed that shift-set RTs were reliably larger than same-set RTs in both the simple shift condition F(1,39) = 28.95, p<.001 and in the cross-talk shift condition F(1,39) = 38.96, p<.001, but were statistically equivalent in the cued shift condition F(1,39) = 2.43. There was a significant interaction between condition and task F(2,78) = 5.97, p<.01. Simple effects tests revealed that shape-naming RTs were reliably larger than colour-naming RTs in the simple shift condition F(1,39) = 47.47, p<.001, the cued shift condition F(1,39) = 50.28, p<.001, and the cross-talk shift condition F(1,39) = 6.55, p<.05, though they were markedly attenuated in the latter. There was also a significant interaction between set-shift and task F(1,39) = 11.58, p<.01. Simple effects tests revealed that colour-naming times were substantially larger in shift-set trials than in same-set trials F(1,39) = 60.22, p<.001, but shape-

Simple Shift Condition



Cross-Talk Shift Condition



Cued Shift Condition

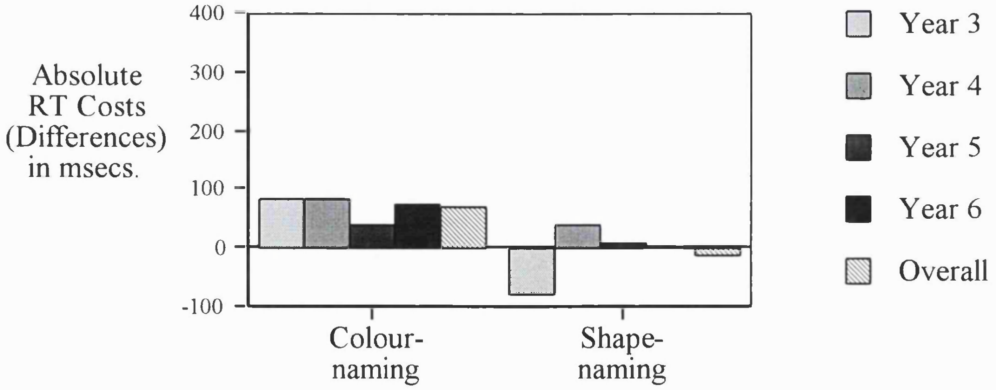


Figure 3.1. Absolute RT Costs by Condition, Naming Task and School Year.

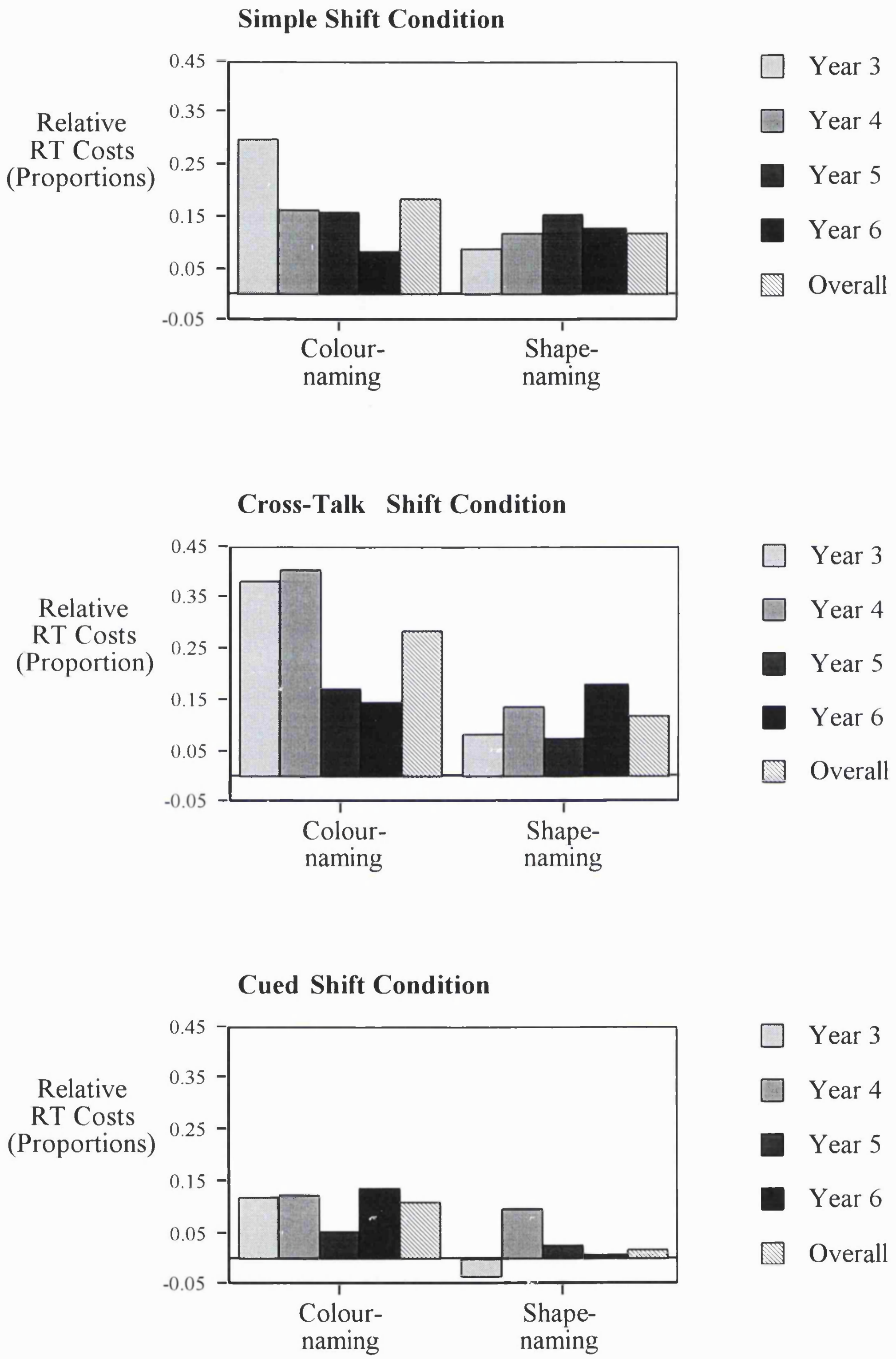


Figure 3.2. Relative RT Costs by Condition, Naming Task and School Year.

From Figure 3.1, it is evident that absolute RT costs decrease with school year for the simple and cross-talk shift conditions, but only on colour-naming trials, and there is no similar change for the cued shift condition. In Figure 3.2 the parallel pattern is discernible for relative RT costs, with the exception of year 3 versus year 4 in the cross-talk shift condition.

To examine whether these associations were reliable, separate GLM procedures were conducted on the absolute and relative RT cost data, for each condition, with the level of α appropriately adjusted to .025. The absolute RT costs in each condition were subjected to a GLM procedure with the factors task and sex, and the co-variates age and ability, which were included in the model. These GLM procedures yielded significant effects only for the simple shift-task condition (block 3). Here, there was a main effect of task $F(1,33) = 8.03$, $p<.01$, confirming that RTs were shorter for the colour-naming task. There was no main effect of sex, but significant interactions between task and age $F(1,33) = 5.83$, $p<.05$, and between task and ability $F(1,33) = 5.86$, $p<.05$. Thus, performance improved with age and an index of ability, but only for the colour-naming task in the simple shift-set condition.

The relative RT costs in each condition were subjected to a GLM procedure with the factors task and sex, and the co-variates age and ability, which were included in the model. These GLM procedures yielded significant effects only for the simple shift-task condition (block 3). Here, there was a main effect of task $F(1,33) = 7.48$, $p<.05$, confirming that RTs were shorter for the colour-naming task. There was no main effect of sex, but significant interactions between task and age $F(1,33) = 5.54$, $p<.05$, and between task and ability $F(1,33) = 5.48$, $p<.05$. Again, performance improved with age and an index of ability, but only for the colour-naming task in the simple shift-set condition.

2) Error Analysis

The means and standard deviations for participant errors, in each block and for the baseline average of colour-naming and shape-naming blocks, are given in Table 3.5. These represent the "difficulty" of each block in terms of the accuracy within a block. Broad parallels with the RT data emerge on inspection. Again, the basic tasks, colour-naming and shape-naming, are not equivalent: shape-naming incurred an average of 1.06 errors per block, while for colour-naming the average was only 0.43. This difference will be taken into account when investigating the effect of shifting between the two tasks.

Unlike for the RT data, error rates for the younger groups do not appear to differ consistently from those of the older children: no clear pattern is evident to visual inspection. For the stop-signal task, failures exceed misses by a clear margin. Relative to a baseline (colour-naming and shape-naming; mean = 0.74) there do appear to be moderate costs to shifting predictably between the two naming tasks (block 3; mean = 1.17), and larger costs when there is stimulus cross-talk (block 4 mean = 3.66). However, costs when the shift was cued by the stimulus itself appear not to differ from the baseline (block 5; mean = 0.69). In the cross-talk shift condition, task error rates exceeded response error rates.

Table 3.5. Mean Errors (counts) and Standard Deviations by Block, Error Type and School Year.

	Colour-naming (block 1)		**Shape-naming (block 2)**		**Baseline (blocks 1 & 2)**		**Sun-Moon (block 6)**	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	0.45	(0.82)	0.91	(1.14)	0.68	(0.81)	6.36	(4.43)
4	0.50	(0.76)	0.75	(0.71)	0.63	(0.69)	4.25	(3.01)
5	0.29	(0.49)	1.14	(1.21)	0.71	(0.81)	4.00	(2.00)
6	0.44	(0.73)	1.44	(1.42)	0.94	(0.95)	4.00	(2.12)
Overall	0.43	(0.70)	1.06	(1.14)	0.74	(0.80)	4.80	(3.25)

	Stop-signal (block 7)			
	Failures		*Misses*	
School Year	Mean	(SD)	Mean	(SD)
3	1.91	(1.64)	0.18	(0.40)
4	2.63	(2.00)	0.13	(0.35)
5	2.57	(1.90)	0.14	(0.38)
6	1.78	(2.11)	0.44	(1.01)
Overall	2.17	(1.85)	0.23	(0.60)

	Simple Shift (block 3)		**Cross-Talk Shift (block 4)**				**Cued Shift (block 5)**	
			Response Errors		*Task Errors*			
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	1.55	(1.29)	2.27	(1.95)	3.09	(2.88)	0.64	(0.92)
4	0.50	(0.76)	0.88	(1.81)	3.00	(0.76)	0.50	(1.07)
5	0.86	(0.69)	0.57	(0.79)	1.29	(1.11)	0.57	(0.79)
6	1.56	(1.01)	0.44	(0.73)	2.33	(1.66)	1.00	(1.12)
Overall	1.17	(1.07)	1.14	(1.63)	2.51	(1.98)	0.69	(0.96)

a) Inhibition Blocks

The proportion of errors per block for the sun-moon task were subjected to a GLM procedure with the factor sex, and the co-variates age and ability. Proportions were used because error rates are binomially distributed and tend to the origin. Since most of these values were equal or close to zero, the arcsin transformation was employed ($y = 2 \text{ arcsine } \sqrt{p}$). The GLM yielded no main effects or significant interactions. The proportion of errors per block for the stop-signal task (failures and misses) were subjected to a repeated measures GLM procedure with the between subjects factor sex, and the co-variates age and ability: the arcsine transformation was employed. The GLM yielded no main effects or significant interactions.

b) Set-shifting Blocks

As in the RT analysis, errors for the set-shifting blocks were divided into same-set and shift-set trials, which were further divided according to naming task. The mean number of errors are summarized for the simple shift condition in Table 3.6, for the cross-talk shift condition in Table 3.7, and for the cued shift condition in Table 3.8. For the cross-talk block, the number of task errors (TE) included in the mean is given in a third column. Again, there is no transparent relationship between school year and error rates for most of the conditions. As observed before though, within the blocks shifting to a different task incurs an error cost, so that shift-set errors tend to exceed same-set error rates. This cost is largest in the cross-talk shift condition, but since the stimulus in this condition varied in two dimensions (colour and shape), and there were two error forms (task and response) this difference must be considered in more detail.

Table 3.6. Mean Errors (counts) and Standard Deviations for the Simple Shift Condition by Same-Set or Shift-Set Trials and School Year.

	Same-Set Trials					
	Both Tasks		*Same Colour*		*Same Shape*	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	0.82	(0.75)	0.45	(0.52)	0.36	(0.50)
4	0.13	(0.35)	0.00	(0.00)	0.13	(0.35)
5	0.14	(0.38)	0.00	(0.00)	0.14	(0.38)
6	0.33	(0.50)	0.00	(0.00)	0.33	(0.50)
Overall	0.40	(0.60)	0.14	(0.36)	0.26	(0.44)

	Shift-Set Trials					
	Both Shifts		*Shift→Colour*		*Shift→Shape*	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	0.73	(0.79)	0.45	(0.52)	0.27	(0.47)
4	0.38	(0.74)	0.00	(0.00)	0.38	(0.74)
5	0.71	(0.76)	0.14	(0.38)	0.57	(0.79)
6	1.22	(0.83)	0.44	(0.53)	0.78	(0.83)
Overall	0.77	(0.81)	0.29	(0.46)	0.49	(0.70

Table 3.7. Mean Errors (counts), Standard Deviations and Task Error Components for the Cross-Talk Shift Condition by Same-Set or Shift-Set Trials and School Year.

	Same-Set Trials								
School Year	*Both Tasks*			*Same Colour*			*Same Shape*		
	Mean	(SD)	TE	Mean	(SD)	TE	Mean	(SD)	TE
3	1.91	(1.81)	0.73	0.64	(0.81)	0.18	1.27	(1.27)	0.55
4	1.25	(1.83)	0.50	0.75	(1.16)	0.25	0.50	(0.76)	0.25
5	0.43	(0.79)	0.29	0.29	(0.76)	0.14	0.14	(0.38)	0.14
6	1.11	(1.05)	0.78	0.44	(0.53)	0.22	0.67	(0.71)	0.56
Overall	1.26	(1.52)	0.60	0.54	(0.82)	0.20	0.71	(0.96)	0.40

	Shift-Set Trials								
School Year	*Both Shifts*			*Shift→Colour*			*Shift→Shape*		
	Mean	(SD)	TE	Mean	(SD)	TE	Mean	(SD)	TE
3	3.45	(1.86)	2.36	1.73	(1.62)	1.00	1.73	(1.68)	1.36
4	2.63	(0.92)	2.50	1.38	(0.74)	1.25	1.25	(1.16)	1.25
5	1.43	(0.79)	1.00	0.14	(0.38)	0.14	1.29	(0.95)	0.86
6	1.67	(1.66)	1.56	0.78	(0.97)	0.67	0.89	(0.93)	0.89
Overall	2.40	(1.63)	1.91	1.09	(1.22)	0.80	1.31	(1.25)	1.11

Table 3.8. Mean Errors (counts) and Standard Deviations for the Cued Shift Condition by Same-Set or Shift-Set Trials and School Year.

	Same-Set Trials					
	Both Tasks		*Same Colour*		*Same Shape*	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	0.18	0.40	0.00	0.00	0.18	0.40
4	0.38	0.74	0.00	0.00	0.38	0.74
5	0.14	0.38	0.14	0.38	0.00	0.00
6	0.11	0.33	0.00	0.00	0.11	0.33
Overall	0.20	0.47	0.03	0.17	0.17	0.45

	Shift-Set Trials					
	Both Shifts		*Shift→Colour*		*Shift→Shape*	
School Year	Mean	(SD)	Mean	(SD)	Mean	(SD)
3	0.45	0.82	0.18	0.40	0.27	0.47
4	0.13	0.35	0.00	0.00	0.13	0.35
5	0.43	0.79	0.29	0.49	0.14	0.38
6	0.89	1.17	0.22	0.44	0.67	1.12
Overall	0.49	0.85	0.17	0.38	0.31	0.68

To investigate whether the observed differences were reliable, the proportion of error trials for each participant was subjected to a repeated measures ANOVA with the factors condition, set-shift and task (as before): the arcsine transformation was employed. This ANOVA yielded main effects for condition F(2,78) = 40.30, p<.001, and set-shift F(1,39) = 10.50, p<.01. Unlike in the RT analysis, however, there was no main effect for task (colour-naming versus shape-naming) F(1,39) = 0.31. Planned comparisons revealed that errors in the cross-talk condition were reliably larger than in the simple shift condition F(1,78) = 43.13, p<.001, and in the cued shift condition F(1,78) = 73.70, p<.001. Unlike in the RT data, the errors in the cued shift condition were not significantly different to those in the simple shift condition F(1,78) = 4.07.

There was no significant interaction between task and set-shift F(2,78) = 0.10, or between task and condition F(2,78) = 2.94, and no three way interaction between the factors F(2,78) = 0.08. However, the main effects for condition and shift were qualified by their interaction F(2,78) = 3.43, p<.05. Simple effects tests revealed that significantly more errors were made in shift-set than same-set trials for both the simple shift condition F(1,39) = 4.48, p<.05, and for the cross-talk shift condition F(1,39) = 9.74, p<.01, but not in the cued shift condition F(1,39) = 1.80.

c) The Error Costs of Set-Shift

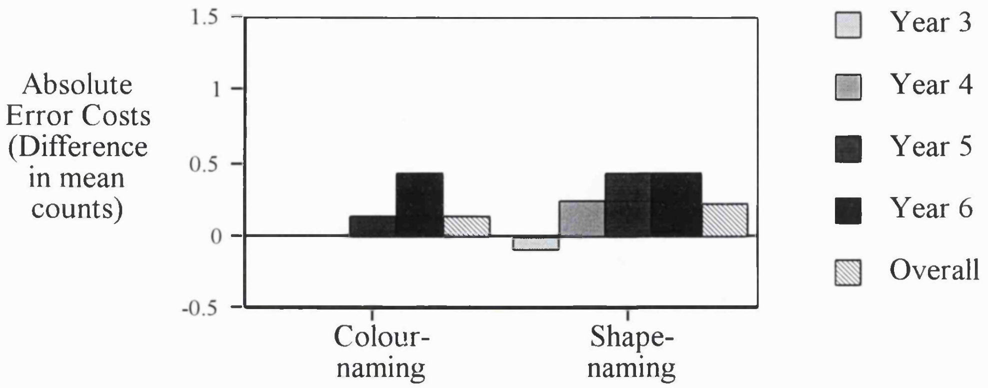
In order to investigate the relationships between these set-shift costs and developmental level or ability, the absolute value error costs of shifting between tasks were calculated for each participant, condition and task using the formula,

error cost = (mean shift-set error – mean same-set error).

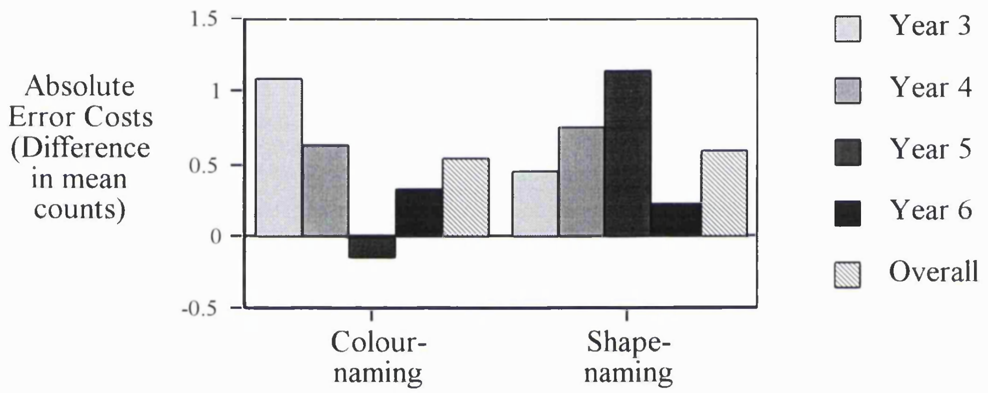
It was not appropriate in this case to calculate the relative error costs since most participant cells were mathematically incalculable. (For example, where same-set errors were nil, the relative cost formula [cost = (shift-set – same-set) / same-set] divides by zero and returns ∞.) The error costs are plotted for each year group and condition in Figure 3.3, while Figure 3.4 distinguishes the response and task error costs for the cross-talk shift condition

Figure 3.3, suggests that error costs do not vary reliably with school year or naming task, but are largest in the cross-talk condition. Again, Figure 3.4 shows that while task error costs exceeded response error costs, no there is no obvious pattern of association with school year or naming task. To examine whether these associations were reliable, the error costs were subjected to separate repeated measures GLM procedures for each condition with the factors task and sex, and the co-variates age and ability. These GLM procedures revealed no main effects or significant interactions for error costs in any condition.

Simple Shift Condition



Cross-Talk Shift Condition



Cued Shift Condition

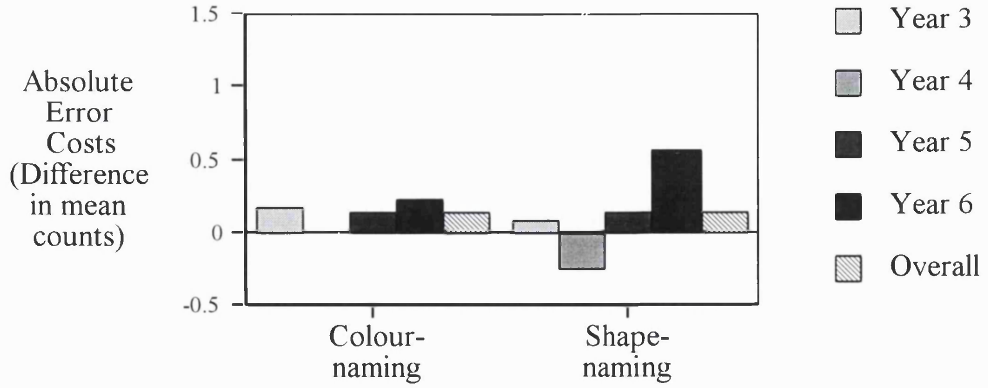
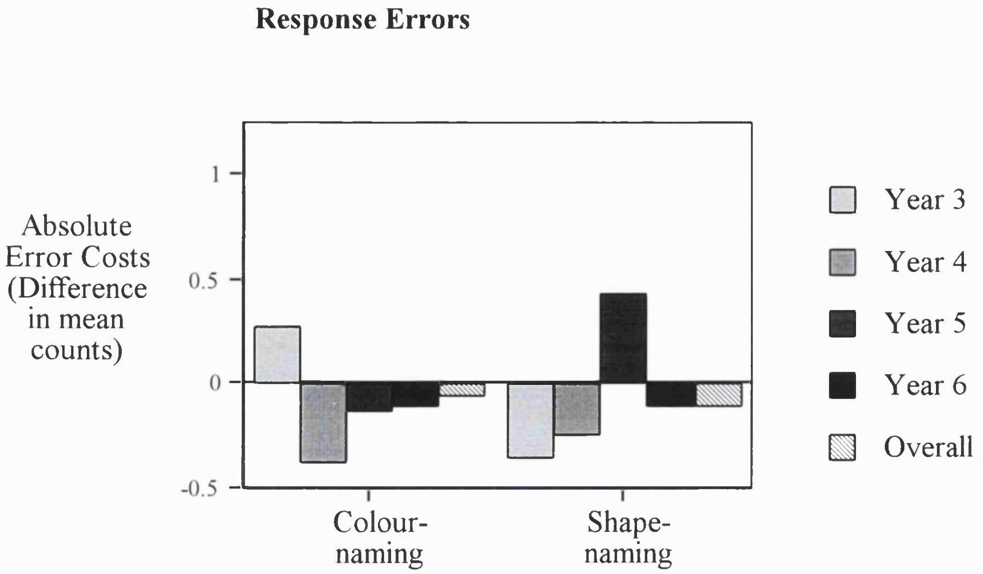


Figure 3.3. Error Costs by Condition, Naming Task and School Year.



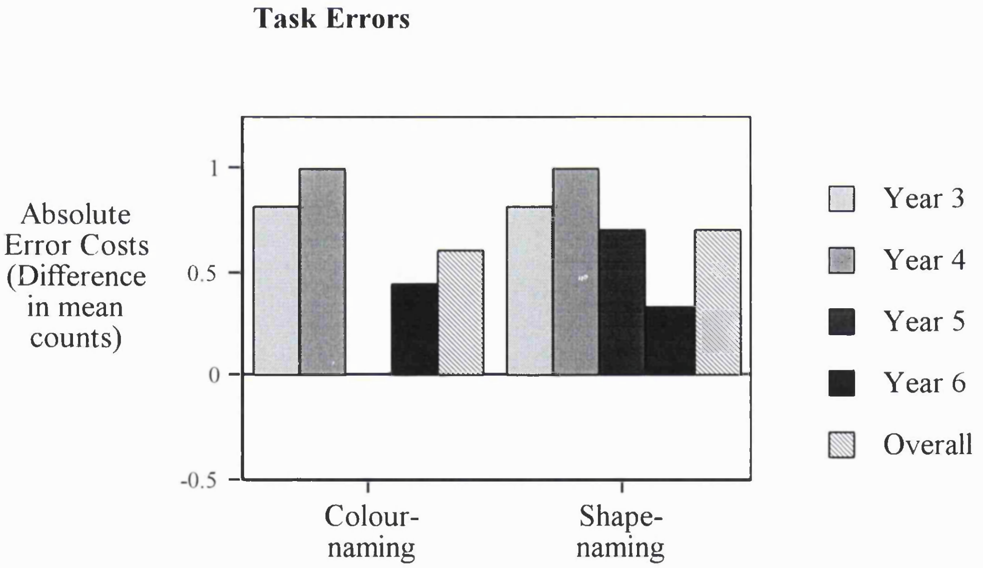


Figure 3.4. Cross-Talk Shift Condition Response and Task Error Costs by Naming Task and School Year.

Discussion

This study confirms that measures of executive functioning can be investigated, using a computer-based battery approach, in children of primary school age. With controlled presentation, and item-by-item measurement of response latencies and error rates, the tasks involved proved appropriate for children as young as 7 years. In the present design, very natural verbal responses were measured to the simplest tasks; colour-naming, shape-naming, and saying "moon" or "sun" where these were the only options. The "material" of interest to the experimenter exists, as it were, "between" the trials and in the speed or accuracy of the response. The procedure is not at all stressful to the participant, and there seems to be no obvious obstacle to using these tests with younger, disordered or delayed children in clinical practice or the cognitive laboratory.

Indeed, the children tested seemed to enjoy the procedure, and were anxious to get from the desk (where they were assessed on the WISC vocabulary and block design sub-tests) to the computer. The atmosphere of novelty around the procedure inspired an impressive degree of co-operation, motivation and concentration from the children. The contribution of these attitudinal components to the process and results of neuropsychological assessment should not be underestimated (Lezak, 1995). Moreover, achieving an index of the participant's "best level" performance is particularly important if results are intended to inform theoretical models of psychological disorder or the assess the progress of an intervention.

Before going on to consider the broad themes that might be extrapolated from the data presented here, I will consider in some detail the methodological limitations of this study: particularly in the sample size and structure, the relationship between RT and error rate, and the fixed order of test presentation. Later, I shall consider how the measures included in the battery (and the relevant constructs) might be validated and the test's reliability assessed, particularly with reference to problems of "task impurity".

1) Methodological Issues

a) The Sample

Since only 40 children were tested for this study, the number of participants in each "cell" of sex and school year is just five, so variability in each cell will be high, and the measures of central tendency and dispersion accordingly very unstable. For an adequate normative data set to be developed, a greater number of children, of both sexes and at each age, would need to be tested. One reason for the small sample size was the selection criteria employed. Teachers were asked to select children in their group who were of "average" ability for their year, not at either extreme of the ability continuum. This was done to ensure that the procedures involved in this pilot were of use with children of this age. However, for proper normative data to be gathered, a full normal sample would have to be tested for each age or school year i.e., one that included children of all abilities, from very low to exceptionally high, provided they were not neuropsychologically impaired or suffering a pervasive developmental psychopathology. Presently, as is clear from the results, the effect of very low or very high ability on the performance of the tasks employed here cannot be determined. Good normative data would need to be collected from several schools, in both urban and rural areas, with a distribution of social class and ethnic origin that approximates closely to that observed in the general British population.

b) The Relationship Between RT and Error rate

The children who participated in this study seemed to have confidence in the idea of working, as they were instructed, "as quickly but accurately as possible," and most expressed in an interest in their mean response times, and number of errors they had made. But this raises the issues of the complicated relationship between RTs and error rates in performance tasks of this kind, and whether children with developmental delay or pervasive developmental psychopathology can work at all in the same way. Part of the motivation for the present investigation was to consider the role of response latencies, as well as error rates, as indices of performance in tests assessing the executive functions. As Posner (1973) demonstrated, response latency is one of the most powerful instruments available to the experimental psychologist for examining the structure or efficiency of information processing systems.

Nevertheless, most conventional neuropsychological procedures, being "paper and pencil" tests, pay little attention to the time taken by participants to complete a task. Where they do, there is usually either a time limit, so that some information is lost, or participant errors are effectively included in the time-score if the participant falters and corrects his or her self, or by the examiner intervening to elaborate upon task instructions. By ignoring the response time elements in their tests, psychologists have accepted the risks of converting fluid measures into discrete ones (e.g., the participant is deemed to have passed or failed a task depending on whether they completed it in the time allocated) and of attending to qualitative differences between groups at the expense of differences in degree.

On issue is that raw error scores are (as in this study) typically counts which have to be transformed in some way to permit full data analysis. But other attendant problems of error rates in single-stimulus designs such as this are also present in the traditional testing methods. For example, errors are frequently the result of momentary lapses in attention, verbal or manual dysfluency, or environmental interruption, which may or may not be obvious to the examiner. Now, an increased vulnerability to such lapses may be a problem of interest in itself, or a component of the test rubric, and should be duly noted (as is particularly the case with executive deficits). However, where the variable under investigation is theoretically independent of these difficulties, errors should be considered separately, and not be punished in the recorded response time.

This jeremiad may be shown to have particular relevance to studies involving children, especially delayed children or those with profound developmental psychopathology. Where distractibility, poor concentration, or inattention are likely to generate more errors than normal, the underlying ability to perform a given test of executive functions such as shifting set or inhibiting responses, could still be examined on those trials on which the participant was successful i.e., in terms of response time considered separately to error frequency. That error rates are rather unstable is often reflected in their low test-retest validity, non-normal distribution, or variances that greatly exceed the mean. Statistical tests therefore return significant effects less frequently for errors, and this may be a fair reflection of their usefulness.

Now in this study, the “typical” performance instructions were given to the children, i.e., the request to work as quickly as possible, but also making as few errors as possible (Sternberg, 1997). Performance therefore may involve a trade-off between speed and accuracy. According to several researchers, (e.g., Wickelgren, 1977; Pachella. 1974) task performance is always governed by a speed-accuracy trade-off function which relates percentage correct responses to the mean RT. The trade-off function is non-linear, non-decreasing (faster performance always means more errors) and will be different across different tasks or conditions. Participants “choose” a position on the trade-off function according to evidence gathered over trials and blocks of trials. However, different task rubric may change the participant’s position on the function. Task instructions that especially value high accuracy by penalising errors, or which emphasise faster responding without explicit concern for error rate, may shift the participant up or down the speed-accuracy trade-off function respectively. Note, though, that RT is less “pure” (in the sense of indexing the processes it is intended to index) if the associated error rate is high: the data may be contaminated by guessing strategies, or responses based on partial stimulus analysis (Sternberg, 1997).

It is not known to what extent normal participants can or do “trade-off” between their speed and accuracy under typical task instructions (Wickelgren, 1977). For example, Pachella (1974) tested participants on the same task three times: first under the typical task instructions, then under a rubric which compelled them to trade for RT at the expense of errors, and finally under the typical instructions again. Pachella showed that participant performance was qualitatively different to before when they returned to the typical task instructions. This suggests that rubrics stressing faster RTs can produce a significant and lasting change in processing strategy. Sternberg (1997) argues that their may be a normative human information processing strategy, distinguishable from those employed under RT or error stress rubrics, that would account for the very direct and orderly RT-error relationships consistently observed across participants in a wide variety of cognitive tasks and experiments.

The question that arises for the present study is whether participants with cognitive impairment, or pervasive developmental psychopathology, adopt the same processing strategies as normally developing children. Other groups or individual children may, for instance, prefer to achieve higher accuracy but at the expense of very long mean RT, or may respond with a competitive mean RT but very low accuracy because of partial information processing. This possibility introduces interpretative difficulties in addition to the failures of task demand (inattention, dysfluency, distraction) outlined above. One might argue convincingly that obsessive psychopathology (e.g., as a feature of autism) would predict a preference for high accuracy and an accordingly laboured RT, while impulsive psychopathology (e.g., as a feature of ADHD) would predict rapid but inaccurate responses.

This is an empirical question, and an interesting one since it applies not only to performance measures of the kind employed in this study, but to all tasks where speed and accuracy are intrinsic to the measure, including therefore most neuropsychological impairments and the associated cognitive tests. Resolving, or more likely accommodating, this issue will entail paying careful attention to both RT and error rates in normal and impaired performance: qualitative differences in apparent *strategy*, between groups and individuals, may prove to be as important for interpretation as the *chronometric* analyses.

c) Order of Task Presentation

One of the main methodological limitations of this study was the fixed order of task presentation: colour naming, shape naming, simple shift, cross-talk shift, cued shift, sun-moon, stop-signal. A fixed presentation order raises several problems for interpreting the available data. Certainly, the order given earlier is not the only one that would have achieved the limited practice needs indicated in the Introduction. Colour and shape naming blocks could have been presented in either order, the two inhibition blocks could have preceded the set shifting blocks, the cued shift block might have been given earlier, and so on. Furthermore, any fixed order of presentation entails the likelihood that performance on a later task is influenced by exposure to an earlier one, but in ways that are not predictable, expected or understood. For example, inhibition processes elicited by the sun-moon task may have facilitated inhibitory processes involved in the subsequent stop-signal task, and there is good reason to prefer, for example, that they were presented in the opposite order.

Similarly, if "set shifting" and "inhibition" processes are not separable control constructs (as Kimberg & Farrah, 1993 have suggested) then earlier practice on set shifting tasks may facilitate or activate inhibitory processes in both the subsequent inhibition tasks. Interpretation of the available data must therefore be qualified by considering the possible effects of a fixed presentation order, and the likely consequences for activation of tasks sets and their sequential interaction. Appropriate ways to investigate the influence of a presentation sequence, and to estimate the costs and benefits of different series will be considered later (see Reliability and Validity below).

2) Preliminary Findings

Bearing in mind the methodological limitations outlined above, some implications of the results available so far, and the suggestions they lead to for future research in executive functions and developmental psychopathology, can be outlined and are summarized below. The inhibition tasks will be discussed before the set shifting tasks, and some broad predictions for the performance of children with profound developmental psychopathology are offered.

a) Inhibition Tasks

i) Sun-Moon Task

This task is surprisingly difficult, but the rubric was among the easiest for the children to grasp. It was without doubt their favourite, and the end of the block was usually marked with a fit of giggles or a loud "phew". Older children were generally faster and made fewer errors than younger groups, but the observed pattern of RT and error means showed most improvement over years 3 to 4 (ages seven to nine). RT performance improved reliably with age and was significantly associated with measured ability.

These results are comparable to those obtained by Gerstadt et al., on their Day-Night task, though they found the greatest performance improvement between ages 4 and 5. The present version was ostensibly simpler, but the more direct opponence between "sun" and "moon" would presumably have increased demand on the inhibitory processes. Accordingly, performance may not reach the putative asymptote until the children are a little older. This version of the cognitive inhibition test might be expected to elicit larger effects in even younger or disordered samples, and thus have greater discriminatory power.

Two qualifications to this interpretation are obvious. First, that the youngest participants in this study had on average lower ability scores than older groups suggests that the age and ability variates are interacting in this task. Second, if the Sun-Moon task is somewhat harder to perform than the Day-Night task, then it might be too difficult for younger children or delayed samples to perform at all, or their performance may fall to an unacceptably low level. Further research on normally developing samples of all ages, including older children, is required to determine the ceiling and floor performance levels on this task, as well as the age- and ability-related normative criteria.

The sun-moon task has yet to employed in sample of delayed children or children with a profound developmental psychopathology, and there are no direct parallels to this task in the literature from which to extend useful predictions regarding the performance of children with profound developmental psychopathology. If one considers the most studied disorders, autistic disorder, ADHD and Tourette syndrome, the possibilities include that all groups would show impaired inhibition (measurable in terms of RT or error rate), that none of them would, or most likely there would be differential impairments. If pressed to conjure a contrast, the existing evidence would suggest that non-delayed children with autism could be quite successful on this cognitive inhibition task, while children with

Tourette syndrome would show poorer performance, and children with ADHD show the most significant impairment. Future research should qualify these guesses.

ii) Stop-Signal Task

This task elicited two very different attitudes from the children. Some thought it was very easy, other claimed it was quite difficult. All were able to explain that, "you have to wait to see if the cross comes." Overall, response latencies and error rates showed no reliable association with age or ability. Also, though on average there were more failures (saying the colour when a stop-signal had been presented) than misses (not responding when there was no signal), this was not statistically significant. Of course, for the present purposes, these results are very positive. They demonstrate that the onset of a stop-signal can be accurately calibrated for individuals (and were in this study) to obviate the effect of differences in processing time. These differences would presumably be age- or ability-related.

The results are compatible with Logan's (1994) own finding, for children aged between seven and eleven on a different version, that the capacity to inhibit responding is relatively unaffected by age. The present version confirms that this is true (over a brief developmental period) using a task in which the primary and stop-signal stimulus are integrated in the same modality. It remains for future research to determine whether this is the case for younger samples, and to stabilize the normative data set. From preliminary analysis, it seems possible that a larger sample (from the same age range as that examined here) might well introduce this paradigm to some subtle developmental trends.

As discussed previously, performance on the stop-signal task has been examined in autistic disorder children by Ozonoff and Strayer (1997). These authors found that autistic disorder children were not impaired, relative to matched controls, on the task. If a parallel can be drawn with the related Go-NoGo paradigm, then it might be predicted that non-delayed children with autism would be relatively successful on this motor inhibition task, while children with Tourette syndrome should show the poorest performance. Children with ADHD might be predicted to show a variable or perhaps intermediate pattern of impairment. Again, only future research will qualify these tentative predictions.

b) Set-Shifting Tasks

i) The effect of different cognitive tasks

Though the basic cognitive tasks of colour-naming and shape-naming were chosen for this study on the assumption that they would have similar processing properties, they clearly are not equivalent in terms of response times or error rates. Shape-naming takes longer and is more error-prone than colour-naming. This difference was necessarily taken into account when investigating the effect of shifting between the two tasks, and may be seen to have important consequences. The disparate effects of cognitive tasks on shifting between them has troubled other researchers in this field. Rogers and Monsell (1995) noted that there was distinct asymmetry between the costs of shifting to a colour-naming task versus shifting to a word-reading task in a Stroop-like shift-set paradigm. The costs of shifting to word-naming were substantial, but there was little cost to performance of shifting to colour-naming; Allport, Styles and Hsieh (1993) also noticed this feature of their data.

This asymmetry is not in the direction one would expect, however. Word-naming is the "stronger" task-set, so shifting to it should be easier. Rogers and

Monsell show that making the stimulus word less legible, and therefore word-naming more difficult, tends to equate the costs of shifting. Thus, the costs of a shift vary inversely with the relative processing demands of the task to be *shifted to*. In contrast, Allport et al., suggest that the non-dominant task-set, colour-naming in the Stroop paradigm, has to attract a heavier "attentional weight", and is therefore more difficult to *disengage from*. In the present study, colour-naming could be considered the "stronger" task, while the shape-naming task was probably more "difficult", yet shifting to colour-naming incurred the greatest RT and error rate costs. The interpretation offered by Allport et al., which depends on the task-set to be shifted from, rather towards, therefore receives some support in the present study.

Of course, in any shift-set paradigm, the "to task" and "from task" arguments are intimately related, since shifting to one task-set will always involve shifting away from another, at least in the sense of which one is currently active or which is inhibited. Therefore, the constructs of "inhibition" and "set-shift" might more usefully be considered as two sides of the same coin, in the same way that "learning" and "memory" are: each serves to complete and express the other. It is assuredly no coincidence that inhibition and set-shifting tasks, despite surface differences, have both been strongly associated with executive functioning. Certainly, any theoretical advance in this arena will depend on a fuller exposition of their inter-dependence.

ii) The effects of endogenous and exogenous control

Relative to a baseline average for colour-naming and shape-naming, there are clear costs to overall (block by block) RT and error rates of shifting between two tasks. These costs appear moderate when the participant is shifting every two trials between unambiguous stimuli (simple shift condition), are somewhat larger when the shift is unpredictably cued by the stimulus itself (cued shift condition), and appear substantial when the stimulus presented affords a response to either task (cross-talk shift condition) even though task order is predictable. For response latencies, these differences between the blocks were statistically reliable. That the costs to performance vary between the three blocks suggests they cannot be attributed merely to the difficulty of keeping two task-sets active at the same time. The alternating runs paradigm is therefore effective in distinguishing the costs of shifting between two task-sets from the cost of maintaining two task-sets in mind.

In interpreting the means for each block, it is evident that there are different effects on performance of the relative contributions of endogenous (internal) and exogenous (external) control over responding. When participants can predict the next set-shift, and prepare to (or begin) configuring or activating it, the costs are smallest. When the subject cannot configure or activate the appropriate set in advance, but an unambiguous stimulus fully cues the set itself, costs appear marginally larger. When the participant can prepare the set in advance, but a stimulus is presented that cues both the appropriate and the now irrelevant task-set (stimulus cross-talk), the costs appear greatest. Monsell's (1995) view that set-shifting involves both a time-limited internal component, and a stimulus-driven completion component, is supported in this study.

One aspect of executive functioning that is not explicitly examined in the present design is that involved in *planning*, as assessed in the Tower of London task for example. (Another is *working memory*, but in definitions of this construct the feature of interest is a personified "executive" corresponding to the domain of executive functions discussed here.) However, planning ability is implicitly addressed in the simple and cross-talk conditions of this battery, since both are assumed to involve the participant actively preparing a task set i.e., anticipating a future state and planning to execute an appropriate response. In the cued shift condition, it is assumed that there is no comparable opportunity to plan the response, or configure the relevant task set. So for the present purposes planning may be considered another aspect of endogenous control.

Notice that in the WCST and the ID-ED shift tests, all stimuli are ambiguous with respect to task-set. The WCST card designs all vary along the dimensions of colour, shape and number, while the ID-ED target array always includes a pink figure and a white line. Therefore both tests confound exogenous and endogenous control in each trial or event and are cross-talk saturated. Moreover, since both tests involve a rule deduction procedure, according to "randomly" changing rules, neither can distinguish cued from predictable shifting within the test session. The present design contrasts the effects of exogenous and endogenous control on performance by manipulating task predictability and stimulus ambiguity: their different effects, just on the performance of normally developing children, is indicative of their importance to a proper analysis of the executive functions.

iii) The effect of shifting within a block

The chief advantage of the alternating runs paradigm, as employed in this study, is that it allows us to contrast performance on shift-set trials and same-set trials *within the same block*. Thus, the effect of holding two task-sets active at once, and any global influence of the test rubric or stimulus dimensions, are controlled for over shifting and same task trials. In particular, this allowed the qualities of the cued shift condition to be distinguished from the simple and cross-talk blocks. Generally within the blocks, trials involving a shift to a different task-set took longer and were more error-prone than trials involving the same task as the previous one. The difference in RTs was statistically reliable only for the simple shift and cross-talk shift blocks. As mentioned before, there were different effects of colour-naming and shape naming within the blocks: shape-naming generally took longer and was less accurate than colour-naming. Again, the RT measure of this effect was weakest in the cued shift condition. Otherwise, both colour-naming and shape-naming RTs were longer in shift-set trials than in same-set trials for all blocks.

The subtle difference between the cued-shift block and the other blocks, in terms of shifting task-set and task type, requires some explanation. One good alternative interpretation would be that this test was fundamentally different to the others in that it amounted to a "say what you see" task. There would be a different task-set configuration for this test. For example, in the simple shift block, the configured sets might be as follows, "expect a colour, say the colour" versus "expect a shape, say the shape". For the cued shift blocks, there might be one set only: "what is it? (colour or shape) Say which one it is." In effect, where there is one task-set, no shift of set is required. The cued shift task-set would be more demanding on processing than the simple shift task-set, since it involves a judgement about stimulus type before a decision about the stimulus dimension, and this might explain why this block's RTs are marginally longer than for the simple shift block on average, but shorter than for the cross-talk block.

The interpretation outlined earlier is that the relative pure exogenous control exerted over responding in the cued shift task is more efficient at configuring the appropriate task-set than preparatory configuration processes. Thus there are two task-sets available, "say the colour" versus "say the shape", but the correct one is elicited fully, without help from an endogenous system, and responding is more automatic (i.e., rapid, involuntary, "in parallel", non-conscious, or however else you wish to define automaticity). Responding would be slower than for the simple shift task because all these processes would happen after stimulus onset, but would be less effortful and relatively more accurate. At the moment it is not possible to distinguish these alternatives, but not doing so will not harm the rational for including the cued shift test in a battery, since both interpretations involve the executive choosing between competing elements of component processes.

iv) Differences between the RT and error indices

In this study, only the performance of normally developing children was examined, and on tasks in which they were expected to be competent, so the observed error rates are intrinsically low. It is not surprising, therefore, that few of the statistical procedures to which the error scores were subjected yielded significant results. However, the error data do resemble the RT data in a few circumscribed ways. Relative to a baseline, moderate error costs to shifting predictably between two naming tasks were observed, and these costs were larger when there was stimulus cross-talk: costs when the shift was cued by the stimulus itself did not differ from the baseline. As in the RT data, *within* the blocks, shifting to a different task incurred an error cost relative to same-set trials, and this cost is largest in the cross-talk shift condition. These findings were statistically reliable. Likewise, reliably more errors were made for shape-naming trials than colour-naming trials in both the simple shift condition, and the cross-talk shift condition, though not in the cued shift condition.

However, unlike for the RT data, error rates for the younger groups did not appear to differ consistently from those of the older children in most of the tests, and there was no reliable difference in error rates between the colour-naming and shape-naming tasks-sets overall. Task error costs exceeded response error costs in the cross-talk shift condition by a wide margin, and this is probably the main reason error costs are largest overall for this block. In summary, it may be seen that error rates do not necessarily parallel all effects observed in the RT data, and may qualify them. Attention is usefully paid to the source of errors and their meaning: in this study, for example, error rates appear more variable within the school year groups than across them.

v) The effects of sex, age and ability

Participant numbers in this study were too small for usual regression analyses, and the GLM procedures employed accordingly produced few significant associations. However, some general developmental trends were also observed in the RT and error data. First, though, note that no effect of sex was detected in this study, for response time or error rates in any of the tasks and conditions. This is perhaps to be expected in a small sample. Also, sex differences are typically found in tasks explicitly assessing the "academic" skills of literacy or mathematical ability, and there were no sex differences in the present sample for vocabulary or block design scores and hence "ability". The girls and boys were also well-matched for age. In addition, sex differences on most tasks from the cognitive laboratory are not usually detected until later in life, and for some tasks they are never identifiable (MacLeod, 1991).

For our purposes, the apparent absence of sex differences may make compiling and expanding the normative data easier, but there is no reason not to suspect that sex differences might be identified in a larger sample, or in groups of younger children. In contrast, some developmental psychopathologies show clear sex differences in incidence (Pennington & Ozonoff, 1996). Autistic disorder and ADHD are both more commonly diagnosed in males than females, though Tourette syndrome has a more equal sex ratio. One consequence of this is that studies of groups of girls with ADHD or autism are rare in the literature, and little is known about potential sex differences in the symptomatology of these disorders.

As regards age, overall set-shift conditions response latency and error rate means were greater for younger children than the older groups. However, as previously noted, it is by no means clear whether this is a function of set-shifting costs or more directly of the basic task processing time. In fact, age-related changes in performance are a feature of the non-shift blocks in this study (colour-naming, shape-naming and the sun-moon task) as they generally are in the cognitive tasks (e.g., the classic Stroop task; McLeod, 1991). The costs of shifting were therefore examined both in absolute and relative (proportional) terms in this study.

Absolute RT costs decreased with school year for both the simple and cross-talk shift conditions, but only on colour-naming trials; and there is no parallel trend in the cued shift condition. The same pattern is discernible for relative RT costs. The only reliable association found for ability was with colour-naming performance in the simple shift-set condition. It was noted in the Method section, however, that age and ability were confounded, by a selection bias, in this sample and were strongly correlated. Nevertheless, the finding of no influence of ability on these tasks is theoretically beneficial for this study, since the executive functions it assesses may therefore be assumed to be relatively independent of "general intelligence" and abilities in the other cognitive domains (e.g., of perception, attention, memory, association and language).

c) Predictions for the Profound Developmental Psychopathologies

In the present study, as noted earlier, the effects of exogenous and endogenous control on performance were contrasted by manipulating task predictability and stimulus ambiguity across shift-set blocks. For two of these blocks, shift-set and same-set trials were compared within the block using an alternating runs design. Accordingly, it is difficult to make clear predictions about the performance of children with profound developmental psychopathology on these kinds of test, since the procedures employed are not directly comparable with those used in previous studies. However, since one of the main motivations behind this study design was to contrast children's performance over the different tasks within the battery, some preliminary hypotheses can be sketched at this point.

i) Tourette Syndrome

For children with "tic and motor" disorders such Tourette syndrome or Sydenham's chorea, executive function deficits as measured on set-shift tasks are not regarded as aspects of the core psychopathology. Rather, inhibitory processes are impaired in these disorders, and no specific prediction will be made concerning the simple, cross-talk, and cued shift tests from the battery. It is only fair to comment, though, that these disorders are associated with deficits in the processing of complex visual information, and in the praxis of responding. These deficits would certainly affect response times and error rates on the set-shift tasks, particularly the cross-talk condition, but not in a way that systematically reflects executive dysfunction.

ii) Autistic Disorder

In the existing literature, children with autistic disorder consistently show impairments in the flexibility of set-shifting, rather than in shifting per se, and appear particularly vulnerable to the capture of control by pre-potent responses or well-learned task-sets (Hughes et al., 1994). These impairments map very directly onto the kinds of behavioural disturbance characteristic of the disorder, including the preference for environmental uniformity, perseverative behaviours, and the seeming inability to disconnect from immediate contact with the environment. Children with autistic disorder would therefore be expected to show different patterns of performance on the simple, cross-talk and cued shift blocks.

Performance would presumably be moderately impaired on the simple shift test, but within the block difficulty should be limited to the shift-set trials which are, to use the terms from Hughes et al., extra-dimensional shifts (while the same-set trials in this study are, by analogy, intra-dimensional shifts). Children with autistic disorder should show significant impairment on the cross-talk shift test, where shift between tasks is combined with ambiguous stimuli that activate both available task sets, and there is conflict between two pre-potent responses. Greatest impairment might be expected for the cued shift test, however, where the key to success is highly flexible shift, and the shift between task sets is unpredictable.

iii) ADHD

Children with ADHD have certainly been shown to exhibit executive deficits in tasks involving set-shifts (as measured by the WCST and Trail Making Test, for example) but the Tower of London planning task appears to be especially sensitive to this disorder (Pennington & Ozonoff, 1996). In contrast to autistic disorder, the core cognitive deficits in ADHD are assumed to be in planning and "sticking to" the appropriate response. Thus, the relevant executive deficits are in configuring and maintaining appropriate task sets. This model is entirely compatible with the cardinal behavioural disturbances observed in ADHD of hyperactivity, impulsivity and distractibility. Accordingly, children with ADHD should be most impaired on set-shifting tests that involve some opportunity for preparation of an appropriate task-set.

Performance on the present battery would be poorest in ADHD for the simple and cross-talk shift conditions, where the shift is predictable. That is, since planning is inefficient, more of the task-set configuration processes will be captured in the response latency, and their difficulty reflected in the error rates. Again, performance would be most impaired on the cross-talk shift condition, where the stimuli are ambiguous with respect to task-set, and do not assist endogenous processes in configuring the relevant one. At this point, the possibly controversial prediction may be made that children with ADHD would be at an *advantage* (relative to children with other disorders and perhaps even normally developing controls) on the cued shift test. In the absence of the opportunity to plan a response, vulnerability to exogenous capture of control might confer a benefit to performance for ADHD children in this shift-set condition. At least, their performance on this block should not be significantly impaired relative to the simple shift task block.

3) Test Development

A good test of psychological processes is one which samples a certain behaviour, under standardised conditions, and includes explicit procedures for obtaining qualitative information about the participant being tested (Murphy & Davidshofer, 1995). The process of test development is a long and technical one, beginning with test construction (according to operational definition in the domain of interest), standardisation of administration procedures and normative data, and concurrent investigation of the test's reliability and validity. The results reported for the each of the five tests employed in this battery (Tables 3.1 and 3.5 in particular) could constitute a preliminary data set for these tasks in the age range studied. However, in light of the methodological criticisms levelled above, and the necessary processes of test development, several further investigations and analyses will be needed to standardise the procedures, and investigate their reliability and validity as tests of executive functioning.

i) Standardisation

The data presented in this study are obviously based on a very small population sample of a limited age range, artificially weighted towards children of "average" ability, and clearly need to be substantially elaborated. As already noted, the tasks employed here need to be investigated in younger and older samples, from around 3 ½ to 16 years of age, and in normally developing children of all abilities. If they are to be useful in the investigation of pervasive developmental psychopathology, a comparison sample of profoundly delayed or low ability children (e.g., with diagnosis of Down's Syndrome or mild to moderate learning difficulty) would serve profitably as an ability matched control.

Compared, for example, to projective techniques, self-report questionnaires or verbally administered tests, computer-based assessment procedures are at an advantage in terms of standardised presentation (Ozonoff, 1995). Obviously, there will be individual and environmental influences on measurement error, as well as the effect of different examiners on participant motivation and comprehension. However, the kind, order and rate of item presentation in computer-based procedures is tightly controlled, and the extraction of qualitative information can be made highly explicit and is relatively independent of interpretation biases.

As indicated earlier, the main problem for the present design is in choosing the (fixed) order of task presentation. First, the influence of previous experience with one task on performance of a later one should be examined by comparing performance on one presentation order with other preferred orders, or with the "standard" experimental sequence of a balanced Latin Square (where no block follows another more than once in a given number of participants). This may seem a huge task, but it might suffice to demonstrate between block parameters in groups of children at, for example, three age levels (e.g., 5 years, 8 years, 14 years). Alternatively, a large age and ability spectrum sample could be tested according to a balanced Latin Square allocation procedure, and the order effect estimated by including "Order" as a factor in a GLM procedure or multivariate ANOVA. For the present design, a balanced Latin Square would produce seven possible presentation sequences. A subset of these might usefully be selected for examination e.g., those which keep each of the "set shift" and "inhibition" tasks together, or those in which the colour and shape naming blocks precede the remaining set shifting tasks.

ii) Reliability

The reliability of a measure is the consistency of its scores in the individual and over time and through space. The first key to reliability is in the standardisation of test administration procedures, and as suggested earlier computer-based assessment procedures are at an advantage in this respect. Another basic principle of reliability is that the greater the number of "items" or "observations" the higher the reliability of the overall "score" will be. Performance measures are also at an advantage in this respect since the identical process (e.g., naming a colour after naming a shape) is examined several times in a block. But this raises a problem particular to executive function measures: the greater the number of stimulus presentations, the more practice there will be in a block, and the less the measure will depend on controlled processing (Duncan, 1995). The length of each block of trials will therefore represent a negotiation, between the need for multiple observations and the need for controlled processing in a novel procedure. In practice, the length of blocks in tasks such as those presented here will also be limited by the putative attentional and concentration deficits of children, especially children with developmental psychopathology or delay.

As discussed earlier, the problem of practice is a significant one for executive function assessment. The main experimental methods devised for investigating the reliability of tests involve repeated presentation: either of the identical test again in test-retest reliability, or of a parallel or alternate form of the test. According to some authors, the test-retest reliability of executive function measures is intrinsically low (Rabbitt, 1997), apparently for good reasons to do with novelty, difficulty and the effects of practice: a test with high test-retest reliability might be judged a poor measure of executive functions for that very reason. Similarly, it may not be possible to construct alternate forms of executive function measures since both form and content should be different across the tasks (Phillips, 1997).

However, it may be possible to construct alternate versions of the tasks used in this study without damaging the novelty and difficulty criteria. If one assumes that task sets and working memories are intensely task-specific (as demanded by connectionist theorists) then tasks involving disparate basic processes to those originally employed may still be examined in terms of their set shifting or inhibition properties. Control over the shift between colour and shape naming might not generalise to the shift between, for example, size or number or position naming in parallel forms. Thus an alternate form of the present tasks could be developed using the basic processes of number and position identification with stimuli such as one or two dots appearing in the left or right portion of a given box. Similarly, the sun-moon task could become the cat-dog test, or the table-chair test or involve any similar pair of well-learned opponent names. The stop signal task might be presented in a "name the shape" format, "unless you see a white dot in the middle".

The problems in constructing alternate forms of the present tasks are both theoretical and practical. Theoretically, if control is in fact achieved by a central executive mechanism, or several higher order controllers, then practice on original tasks will generalise to the alternate forms. Practically, as has been shown in discussing the different effects of shifting between two tasks (i.e., shifting from colour to shape naming appears qualitatively different to the reverse shift) the processing components of different tasks do not directly predict their relationship to control. In the short term, it might be sufficient to construct "best guess" alternate forms, and compare their properties carefully in a large enough sample, re-testing with the same normally developing participants after a suitable delay.

Furthermore, task impurity remains a problem for performances indices of this kind. Again, since there are different effects of basic cognitive tasks on the efficiency of set shifting and inhibition, then the measure of executive functioning in intensely dependent on the particular cognitive processes involved. The indices derived from the present design are dependent (at least) on intact comprehension of verbal instructions, sustained attention, visual perception, colour cognition, form recognition, colour and form vocabulary, short term memory for the tasks involved, vocal response output, and any of a suite of demand conditions such as "monitoring" and "error correction". Fortunately, these are partially assessed and practised in this design during the first two blocks (basic colour- and shape-naming), and absence of impairment in these domains would have to be soundly demonstrated if executive dysfunction was to be inferred from poor performance on the present tasks. However, the "emergent" properties of task performance observed in even this simple design (i.e., the disparate efficiency in shifting between colour and shape naming sets) raises again the problems of validating the constructs on which the procedures are based.

iii) Validity

The validity of a measure is its "meaningfulness", which may be understood as its relationship to the psychological construct it operationalizes (construct validity), and its predictive power regarding individuals or their behaviour (criterion validity).

According to Phillips (1997) the criterion validity of executive function measures has usually been assessed in terms of sensitivity to frontal lobe lesions, which is clearly problematic for the reasons outlined earlier. However, a broad and systematic relationship between executive functioning and the pre-frontal cortices has been observed in the literature, and the sensitivity of the present tasks to frontal damage would certainly be one of the most important investigations of their validity. It would be useful to show that performance on the present tasks is impaired in at least some individuals with known pre-frontal lesions or a diagnosed "dysexecutive syndrome".

These conditions are relatively rare in children, but a useful validation study could be undertaken with adults patients, providing one assumes that the nature of control processes in children is substantially continuous with that in adults (which they might not be). Strictly speaking, it would not be so useful to attempt to *validate* these tasks on groups of children with pervasive developmental psychopathology: this would entail a circularity of definition and subsequent interpretation of the form, "children with developmental psychopathology perform poorly on task A, and have executive function deficits, so task A is a measure of executive functioning," unless the previous studies are accepted to have demonstrated executive dysfunction in these groups. Of course this depends entirely on the construct validity of previously employed measures and therefore in turn the construct validity of the present tasks.

A related validation procedure is to examine the extent to which the measure of interest correlates with different measures of the same construct (convergent validity). It would certainly be necessary to show that in certain circumstances the tasks presented here correlate with (and are affected by the same impairments as) other tests of executive functioning. One way to examine this would be to select a large sample of older, normally developing children and investigate the relationship between their performance on the present tasks and the WCST, a classic Stroop test, and a test not designed to measure executive functions such as Vocabulary or Recognition Memory. The present tasks should correlate better with the tests of executive functioning. Obvious problems with this approach include that in normal samples, the convergent validities of neuropsychological tests tend not to differ significantly from their *divergent* validities, even when the effect of general abilities are partialled out (Burgess, 1997) and that such a the procedure implicitly assumes the convergent and construct validity of other executive tests (which I have already gone some way to dispute).

Construct validity is highly prized, indeed some authors have advanced the argument that construct validity *is the* validity (e.g., Cronbach, 1988), while others have argued less technically that, "a neuropsychological measure is only as good as the theory on which it is based," (Davies, 1996). It is disappointing therefore that the validity of core executive function constructs (set shifting, inhibition, planning etc.,) is so poor, and the present study therefore also suffers from the paucity of experimental research to support these terms. As demonstrated in the present results, "set shifting", "inhibition", "working memory" and "planning" are integrated and perhaps inseparable aspects of appropriate task performance. Thus, Kimberg and Farrah (1993) can show that such functions can be achieved in structurally identical models. These constructs may represent only descriptive aspects of particular task demands, and may not map at all adequately onto detailed specifications of task performance (Rabbitt, 1997).

However, several arguments may be levelled in favour of the approach adopted here. First, "inhibition" is an under-researched attentional property of cognitive systems, which has only recently attracted the same interest as its correlate "selection". The clever experiments by Tipper and colleagues on negative priming (Tipper, 1985; Tipper & Driver, 1988; Allport, Tipper & Chmiel, 1985), are rare examples of direct investigation and interpretation of inhibitory processes. Similarly, experimental investigations of task set shift are not common in the literature, and only two groups have made recent contributions to the area (Allport, Styles & Hsieh, 1994; Rogers & Monsell, 1995). It may be hoped that further investigation of these (related) processes will elucidate their nature and function.

Second, the task set approach, which partials performance into separate task systems, task set configuration and task set shifts, represents an important advance on "attentional" interpretations. It is compatible with successful connectionist approaches to cognitive modelling, and the incorporated constructs are therefore testable in more limited and specific ways. The ABAB paradigm, for example, allows a more fine-grained analysis of task shift costs, without involving rule-deduction processes or the configuration of entirely novel or difficult stimulus-response sets mappings. There is less task impurity, and what there is may be made more explicit. It is clear, however, that a good deal of further work needs to be done.

4) Conclusion

The control of cognition and action is no longer the mystery it was to James. The classical critique of the "supervisory" or "executive" system as a processing homunculus (little man inside the head) has recently been addressed in experimental paradigms that attempt to distinguish task performance from task-set configuration. Under the rubric of executive or frontal lobe functions, several relatively independent systems have been identifiable. Modularity within the executive functions has been mapped onto explanations of some of the profound developmental psychopathologies.

In the present study, some of these newer experimental paradigms have been shown to be potentially useful indices of executive functioning in children. Further work is necessary to expand and qualify this analysis. From the methodological limitations and test development requirements outlined above, several specific investigations are needed. First, data on the present battery design has to be expanded to include a larger population sample, including the full range of intellectual abilities, that matches the social and ethnic make-up of the broader community. Second, the effects of a fixed order of task presentation on performance has to be assessed and accounted for using mixed presentation designs. Third, the criterion validity of the measures should be examined in a population with known pre-frontal damage or dysexecutive syndrome. Fourth, the construct validity of the measure should be addressed by comparison with other executive function and non-executive tests, and explicit modelling in connectionist or task set terms. In the longer term, by comparing the performance of groups or individuals with known or suspected executive dysfunction to normative RTs and error rates, it may be possible to discriminate those components of executive function that are deficient in disparate developmental psychopathologies. Ultimately, a standardised assessment battery for executive function may therefore be developed.

Appendices

1) Statistical Appendix

The General Linear Model (GLM) may require some introduction. Its properties have been succinctly summarized by Howell (1995), who shows that the analysis of variance model (ANOVA) is a special case of the multiple regression model (MR), and both are special cases of the General Linear Model. The General Linear Model takes the form

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

where **X** is a matrix (rows and columns) of the data variables, **b** is a vector (column) of "weights", **y** is a vector of the criterion data, and **e** is a vector of errors.

The MR model is a special case of this equation where **X** is a matrix of the predictor variables, **b** is a vector of regression coefficients, and **y** is a vector of the dependent (predicted) variable. The ANOVA model is a special case of the equation where **X** is a design matrix relating participants to treatments, **b** is a vector of the treatments and their mean, and **y** is vector of the participant means. In the General Linear Model, the "test" is on the two sides of the linear equation. Thematically, for MR the test is "Are the two sides the same?", while for ANOVA the test is, "Are the two sides different?", and these tests are equivalent.

The advantage of the fundamental General Linear Model procedure is that it can deal with variables both as predictors (as in MR) or as levels of factors (as in

ANOVA). The procedure can therefore return indices of the main effects and interactions of co-variates, rather than merely partial them out of the analysis. As in MR and ANOVA, General Linear Model statistics are distributed on the ratio F. One current disadvantage of the GLM procedure is that, being a relatively new-comer to the statistical arena, conventions for describing its outputs have not been established in the literature. In this report, GLM terms and outputs are described using ANOVA conventions.

2) Information Sheet For Children

3) Information Sheet For Parents

4) Consent Form for Children

5) Consent Form for Parents

6) Letter of Ethical Approval

Sub-Department of Clinical Health Psychology
UNIVERSITY COLLEGE LONDON
GOWER STREET LONDON WC1E 6BT

Dr Tony Charman
Lecturer in Psychology
UCL: 0171 380 7777 ext. 5945

Dr Matthew Jones-Chesters
Clinical Psychologist in Training
UCH: 0171 530 5111 ext. 5015

INFORMATION SHEET FOR CHILDREN CONFIDENTIAL

Title of this research study.
A Computer-based Test of Executive Functioning in Children

Researchers: Dr Matthew Jones-Chesters and Dr Tony Charman

About this research study.
This study uses some easy "games", done on computer, to try to find out how children can
plan to do something
change from doing one thing to doing another
or stop themselves doing something.
We would like to see how children get better at doing these things as they get older.
We are asking children of between 4 and 14 years old to take part.

Children will be asked to name colours, name shapes, decide either to name the colour or the shape of a picture, and decide if two pictures are the same or different. We would like to know whether we can use the computer to see how well children do these things.

If you take part in this study.
The headteacher of your school has allowed us to invite some children from your school to help us with this study. If you decide to take part, you will be seen at school at a time that is good for you and your teacher. You will then spend up to 40 minutes, with Matthew Jones-Chesters, naming pictures on the computer and answering some questions about words you might know. Children usually find that this is fun to do.

You do not have to take part in the study if you do not want to. If you decide to take part, you can stop at any time without giving a reason. All the information gathered in this study will be kept confidential to the researchers, and your name will not be used.

All proposals for research using human subjects are reviewed by an ethics committee before they can proceed. This proposal was reviewed by the Joint UCL/UCLH Committees on the Ethics of Human Research. If you would like more information on the research study please feel free to contact Dr Tony Charman at the Sub-department of Clinical Health Psychology, University College London, Gower Street, London, WC1E 6BT, telephone number 0171 380 7777 extension 5945.

Sub-Department of Clinical Health Psychology
UNIVERSITY COLLEGE LONDON
GOWER STREET LONDON WC1E 6BT

Dr Tony Charman
Lecturer in Psychology
UCL: 0171 380 7777 ext. 5945

Dr Matthew Jones-Chesters
Clinical Psychologist in Training
UCH: 0171 530 5111 ext. 5015

INFORMATION SHEET FOR PARENTS **CONFIDENTIAL**

Title of this research study.
A Computer-based Test of Executive Functioning in Children

Researchers: Dr Matthew Jones-Chesters and Dr Tony Charman

About this research study.
This research study is about using some easy to learn tasks, done on computer, to try to find out more about children's "executive functioning" i.e., the ways children can
plan to do something
change from doing one thing to another
or stop themselves from doing something.
We would like to see how children get better at setting themselves these "mental sets" as they get older, and are inviting children of between 4 and 14 years old to take part.

The tasks we are using on the computer include naming colours, naming shapes, deciding whether to name the colour or the shape of a picture, and deciding whether two pictures are the same or different. We would like to know whether we can use the computer to measure how well children do on these tasks, and whether they are useful measures of "executive functioning".

If you take part in this research study.
The headteacher of your child's school has given us permission to invite some children to help us with this study. If you decide that your child may take part, they will be seen at school at a time convenient to the child and their teacher. Your child will then spend up to 40 minutes, with Matthew Jones-Chesters, doing the tasks on computer and answering some questions that provide an estimate of their general abilities. Children usually find the computer tasks are fun to do, and not at all stressful.

You and your child do not have to take part in the study if you do not want to. If you decide to take part, you can withdraw your child at any time without having to give a reason. All the information gathered in this study will be kept confidential to the researchers, and no names will be used. All proposals for research using human subjects are reviewed by an ethics committee before they can proceed. This proposal was reviewed by the Joint UCL/UCLH Committees on the Ethics of Human Research.

If you would like more information on the research study please feel free to contact Dr Tony Charman at the Sub-department of Clinical Health Psychology, University College London, Gower Street, London, WC1E 6BT, telephone number 0171 380 7777 extension 5945.

Sub-Department of Clinical Health Psychology
UNIVERSITY COLLEGE LONDON
GOWER STREET LONDON WC1E 6BT

CONSENT FORM FOR PARENTS **CONFIDENTIAL**

Title of this research study.
A Computer-based Test of Executive Functioning in Children

Researchers
Dr Matthew Jones-Chesters and Dr Tony Charman

To be completed by the parents: Delete as necessary

1) Have you read the information sheet about this study? YES / NO

2) Have you had the opportunity to ask questions
and discuss this study? YES / NO

3) Have you received satisfactory answers to your questions? YES / NO

4) Have you received enough information about this study YES / NO

5) Who have you spoken to about this study..

6) Do you understand that you are free to withdraw from this study

at any time
without having to give a reason YES / NO

7) Do you agree to take part in this study? YES / NO

Signed..Date......................................

Your name in block letters...

Researcher's name: Matthew Jones-Chesters Signed.............................

Date...............................

Sub-Department of Clinical Health Psychology
UNIVERSITY COLLEGE LONDON
GOWER STREET LONDON WC1E 6BT

CONSENT FORM FOR CHILDREN **CONFIDENTIAL**

Title of this research study.
A Computer-based Test of Executive Functioning in Children

Researchers
Dr Matthew Jones-Chesters and Dr Tony Charman

To be completed by the child: ....Delete as necessary

1) Have you read the information sheet about this study? YES / NO

2) Have you had the opportunity to ask questions and discuss this study? YES / NO

3) Have you received satisfactory answers to your questions? YES / NO

4) Have you received enough information about this study YES / NO

5) Who have you spoken to about this study...

6) Do you understand that you are free to withdraw from this study

at any time
without having to give a reason YES / NO

7) Do you agree to take part in this study? YES / NO

Signed...Date.......................................

Your name in block letters...

Researcher's name: . Matthew Jones-Chesters Signed..............................

Date................................

UCL
HOSPITALS

The University College London Hospitals

The Joint UCL/UCLH Committees on the Ethics of Human Research

Committee A Chairman: Dr F D Thompson

Please address all correspondence to:
Mrs Iwona Nowicka
Research & Development Directorate
9th Floor, St Martin's House
140 Tottenham Court Road, LONDON W1P 9LN
Tel. 0171- 380 9579 Fax 0171-380 9937
e-mail: i.nowicka@academic.uclh.nthames.nhs.uk

Dr MH Jones Chesters
Clinical Psychology in Training
Sub-department of Clinial Health Psychology
UCL
1-19 Torrington Place

10 March 1998

Dear Dr Jones Chesters

Study No: **97/0361**
Title: **A computer-based test of executive functioning in children**

Thank you very much for your letter of the 22nd February and the simplified version of the information sheet. I confirm that this study is approved by Chairman's Action.

Please note that it is important that you notify the Committee of any adverse events or changes (name of investigator etc) relating to this project. You should also notify the Committee on completion of the project, or indeed if the project is abandoned. **Please remember to quote the above number in any correspondence.**

Yours sincerely

I, Nowicka

Iwona J Nowicka
Administrator, UCL/UCLH Ethics Review Committees

cc.
Dr Tony Charman
Lecturer in Psychology
Behavioural Science Unit
Institute of Child Health
30 Guilford Street
London WC1N 1EH

jo10mr/ijn/10 March 1998

References

Allport, A. (1980). Attention and performance. In G. Claxton (Ed.), Cognitive Psychology: New Directions (pp. 26-64). London: Routledge & Kegan Paul.

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 & Performance XIV. Cambridge, MA: MIT Press.

Allport, A., Styles, E. A. & Hsieh, S. (1993). Shifting intentional set: exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), Attention & Performance XV. Cambridge, MA: MIT Press.

Allport, A., Tipper, S. P. & Chmiel, N. J. C. (1985). Perceptual integration and post categorical filtering. In M. I. Posner & O. S. M. Marin (Eds.), Attention & Performance XI. Hillsdale, NJ: Lawrence Erlbaum Associates.

Allport, D. A., Antonis, B. & Reynolds, P. (1972). On the division of attention: a disproof of the single-channel hypothesis. Quarterly Journal of Experimental Psychology, 24, 225-235.

Anderson, S. W., Damasio, H., Jones, R. D. & Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. Journal of Clinical & Experimental Neuropsychology, 13, 909-992.

Armitage, S. G. (1946). Analysis of certain psychological tests used for the evaluation of brain injury. Psychological Monographs, 60, all number 277.

Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), The Psychology of Learning and Motivation: Advances in Research and Theory, 2 (pp. 89-195). New York: Academic Press.

Baddeley, A. (1990). Human Memory: Theory and Practice. London: Lawrence Erlbaum Associates.

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

Baddeley, A., Della Salla, S., Gray, C., Papagno, C. & Spinnler, H. (1997). Testing central executive function with a pencil and paper test. In P. Rabbitt (Ed.), Methodology of Frontal and Executive Function. Hove: Psychology Press.

Barkley, R. A., Grodzinsky, G. & Du Paul, G. J. (1992). Frontal lobe functions in attention deficit disorder with and without hyperactivity: a review and research report. Journal of Abnormal Child Psychology, 20, 163-188.

Baron-Cohen, S., Cross, P., Crowson, M. & Robertson, M. (1994). Can children with Gilles de la Tourette syndrome edit their intentions?. Psychological Medicine, 24, 29-40.

Baron-Cohen, S., Leslie, A. M. & Frith, U. (1985). Does the autistic child have a "theory of mind"? Cognition, 21, 37-46.

Benton, A. L. & Hamsher, K. de. S. (1989). Multilingual Aphasia Examination. Iowa: AJC Associates.

Bornstein, M. H. (1985). On the development of colour naming in young children: data and theory. Brain & Language, 26, 72-93.

Broadbent, D. E. (1958). Perception and Communication. London: Pergamon.

Broadbent, D. E. (1982). Task combination and selective intake of information. Acta Psychologica, 50, 253-290.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), Methodology of Frontal and Executive Function. Hove: Psychology Press.

Casey, B. J., Vauss, Y. C., Chused, A. & Swedo, S. E. (1994). Cognitive functioning in Sydenham's Chorea: II Executive functioning. Developmental Neuropsychology, 10, 89-96.

Cattell, R. B. & Cattell, A. K. S. (1960). The Individual or Group Culture Fair Intelligence Test. Champaign, IL.: IPAT.

Cohen, J. D., Dunbar, K. & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. Psychological Review, 97, 332- 361.

Cronbach, L. J. (1988). Five perspectives on the validity argument. In H. Wainer & H. Brown (Eds.), Test Validity. Hillsdale, NJ: Erlbaum.

Davidoff, J. (1991). Cognition Through Colour. Cambridge, MA: MIT/Bradford.

Davies, S. (1996). Neuropsychological assessment of the older person. In R. Woods (Ed.), Clinical Psychology of Older Adults.

Denny-Brown, D. (1958). The nature of apraxia. Journal of Nervous & Mental Diseases, 126, 9-32.

Diamond, A. & Goldman-Racik, P. S. (1986). Comparative Development in human infants and infant rhesus monkeys of cognitive functions that depend on prefrontal cortex. Society of Neuroscience Abstracts, 12, 742.

Diamond, A. & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do." Developmental Psychobiology, 29, 315-334.

Drewe, E. A. (1975). An experimental investigation of Luria's theory of the effects of frontal lobe lesions in man. Neuropsychologia, 13, 421-429.

Duncan, J. (1995). Attention, intelligence and the frontal lobes. In M. Rutter (Ed), Developmental Neuropsychiatry. New York: Guildford.

Fodor, J. (1983). The Modularity of Mind. Cambridge, MA: MIT Press.

Gedye, A. ((1991). Tourette syndrome attributed to frontal lobe dysfunction: numerous etiologies involved. Journal of Clinical Psychology, 47, 233-252.

Gerstadt, C. L., Hong, Y. J. & Diamond, A. (1994). The relationship between cognition and performance: Performance of children aged between 3 ½ - 7 years old on a Stroop-like day-night task. Cognition, 53, 129-153.

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

Heaton, R. K. (1981). A Manual for the Wisconsin Card Sorting Test. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Chelune, C. J., Talley, J. L., Kay, G. G. & Curtis, G. (1993). Wisconsin Card Sorting Test Manual: Revised and Expanded. Odessa, FL: Psychological Assessment Resources.

Hebb, D. O. (1945). Man's frontal lobes: A critical review. Archives of Neurology & Psychology, 54, 10-24.

Hobson, R. P. (1989). Beyond cognition: a theory of autism. In G. Dawson (Ed.), Autism: Nature, Diagnosis and Treatment. New York: Guildford.

Howell, D. C. (1995). Statistical Methods for Psychology, (4th edition). Belmont, CA: Duxbury Press.

Hughes, C., Russell, J. & Robbins, T. W. (1994). Evidence for executive dysfunction in autism. Neuropsychologia, 32, 477-492.

James, W. (1890). The Principles of Psychology. New York: Dover.

Jersild, A. T. (1927). Mental set and shift. Archives of Psychology, all number 89.

Jones-Gotman, M. & Milner, B. (1977). Design fluency: the invention of nonsense drawings after focal cortical lesions. Neuropsychologia, 15, 653-674.

Kimberg, D. Y. & Farrah, M. J. (1993). A unified account of impairments following frontal lobe damage: the role of working memory in complex organised behaviour. Journal of Experimental Psychology: General, 122, 411-428.

Klein, G. S. (1964). Semantic power measured through the interference of words with colour naming. American Journal of Psychology, 77, 576-588.

Kopelman, M. D. (1991). Frontal dysfunction and memory deficits in the alcoholic Korsakoff syndrome and Alzheimer-type dementia. Brain, 114, 117-137.

Levin, H. S., Goldstein, F. C., Williams, D. H. & Eisenberg, H. M. (1991). The contribution of frontal lobe lesions to the neurobehavioural outcome of closed head injury. In H. S. Levin, H. M. Eisenberg & A. L. Benton, (Eds.), Frontal Lobe Function and Dysfunction. New York: Oxford University Press.

Lezak, M. D. (1995). Neuropsychological Assessment: Third Edition. New York: Oxford University Press.

Lhermitte, F. (1983). Utilization behaviour and its relation to lesions of the frontal lobes. Brain, 106, 237-255.

Logan, G. (1994). On the ability to inhibit thought and action: a user's guide to the stop-signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), Inhibitory Processes in Attention, Memory and Language. San Diego: Academic Press.

Logan, G. D. (1988). Towards an instance theory of automatization. Psychological Review, 95, 492-527.

MacLeod, C. (1991). Half a century of research on the Stroop effect: an integrative review. Psychological Bulletin, 109, 163-203.

Milner, B. (1964). Some effects of frontal lobectomy in man. In J. Warren & K. Akert (Eds.). The Frontal Granular Cortex and Behavior. New York: McGraw Hill.

Moffitt, T. E. (1993). The neuropsychology of conduct disorder. Development & Psychopathology, 5, 135-151.

Monsell, S. (1995). Control of mental processes. In V. Bruce (Ed.), Great Unsolved Mysteries of Mind. Hove: Lawrence Earlbaum Associates.

Monsell, S., Taylor, T. & Murphy, K. (in preparation). Stroop interference as task-set interference: evidence from effects of lexicality and word frequency.

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

Newell, A. (1980). Reasoning, problem-solving and decision processes. In R. Nickerson (Ed.), Attention & Performance VIII. Hillsdale, NJ: Lawrence Earlbaum Associates.

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, 4. New York: Plenum.

Ozonoff, S & McEvoy, R. E. (1994). A longitudinal study of executive function and theory of mind in autism. Development & Psychopathology, 6, 415-431.

Ozonoff, S. (1995). Reliability and validity of the Wisconsin Card Sorting Test in studies of autism. Neuropsychology, 9, 491-500.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), Human Information Processing: Tutorials in Performance and Cognition. Hillsdale, NJ: Earlbaum.

Pardo, J. V., Pardo, P. J., Janer, K. W. & Reichle, M. E. (1990). The anterior cingulate cortex mediates processing in the Stroop attentional conflict paradigm. Proceedings of the National Academy of Sciences, 87, 256-259.

Pashler, H. (1993). Dual task interference and elementary mental mechanisms. In D. E. Meyer & S. Kornblum (Eds.), Attention & Performance XIV. Cambridge, MA: MIT Press.

Pennington, B. F. & Ozonoff S. (1996). Executive functions and developmental psychopathology. Journal of Child Psychology & Psychiatry, 37, 51-87.

Phillips, L. H. (1997). Do frontal tests measure executive function? Issues of assessment and evidence from fluency tests. In P. Rabbitt (Ed.), Methodology of Frontal and Executive Function. Hove: Psychology Press.

Porteus, S. D. (1965). Porteus Maze Test: Fifty Year's Application. New York: Psychological Corporation.

Posner, M. I. (1973). Chronometric Explorations of Mind. New York: Basic Books.

Rabbitt, P. (1997). Introduction: methodology and models in the study of executive function. In P. Rabbitt (Ed.), Methodology of Frontal and Executive Function. Hove: Psychology Press.

Raven, J. C. (1960). Guide to the Standard Progressive Matrices. London: H. K. Lewis.

Reason, J. (1990). Human Error. Cambridge: Cambridge University Press.

Reitan, R. M. & Wolfson, D. (1994). A selective and critical review of neuropsychological deficits and the frontal lobes. Neuropsychology Review, 4, 161-198.

Rogers, R. D. & Monsell, S. (1995). The costs of a predictable switch between simple cognitive tasks. Journal of Experimental Psychology: General, 124, 207-231.

Schneider, W. & Detweiler, M. (1988). The role of practice in dual task performance: towards workload modelling in a connectionist/control architecture. Human Factors, 30, 539-566.

Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing, I: detection, search and attention. Psychological Review, 84, 1-66.

Schneider, W. (1993). Varieties of working memory as seen in biology and in connectionist/control architectures. Memory & Cognition, 21, 184-192.

Segalowitz, S. J., Unsal, A. & Dywan, J. (1992). CNV evidence for the distinctiveness of frontal and posterior neural processes in a traumatic brain-injured population. Journal of Clinical & Experimental Neuropsychology, 14, 545-565.

Shallice, T. & Burgess, P. W. (1991). Higher order cognitive impairments and frontal lobe lesions in man. In H. S. Levin, H. M. Eisenberg & A. L. Benton (Eds.), Frontal Lobe Function and Dysfunction. New York: Oxford University Press.

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

Shallice, T. (1988). From Neuropsychology to Mental Structure. Cambridge: Cambridge University Press.

Shallice, T., Burgess, P. W., Schon, F. & Baxter, D. M. (1989). The origins of utilization behaviour. Brain, 112, 1587-1598.

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing II: Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Spector, A. & Biederman, I. (1976). Mental set and mental shift revisited. American Journal of Psychology, 89, 669-679.

Sternberg, R. J. (1997). How we compare objects. In D. Scarborough & R. J. Sternberg (Eds.), Invitation to Cognitive Science IV: Methods, Models and Conceptual Issues. Cambridge MA: MIT Press.

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

Stuss, D. T. (1992). Biological and psychological development of executive functions. Brain & Cognition, 20, 8-23.

Temple, C. M., Carney, R. A. & Mullarkey, S. (1996). Frontal lobe function and executive skills in children with Turner's syndrome. Developmental Neuropsychology, 12, 343-363.

Tipper, S. P. & Driver, J. (1988). Negative priming between pictures and words: evidence for semantic analysis of ignored stimuli. Memory & Cognition, 16, 64-70.

Tipper, S. P. (1985). The negative priming effect: inhibitory effects of ignored primes. Quarterly Journal of Experimental Psychology, 37, 571-590.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. Quarterly Journal of Experimental Psychology, 37, 571-590.

Wechsler, D. (1981). WAIS-R Manual. New York: Psychological Corporation.

Wechsler, D. (1990). Wechsler Pre-school and Primary Scale of Intelligence Revised UK Edition. Kent: The Psychological Corporation.

Wechsler, D. (1991). Wechsler Intelligence Scale for Children Third Edition UK. Kent: The Psychological Corporation.

Weiskrantz, L (1992). Introduction: Disassociated issues. In A. D. Milner & M. D. Rugg (Eds.), The Neuropsychology of Consciousness. London: Academic Press.

Welsh, M. C. & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. Developmental Psychology, 4, 199-230.

Welsh, M. C., Pennington, B. F., Ozonoff, S. & Rouse, B. (1990). Neuropsychology of early-treated phenylketonuria: Specific executive function deficits. Child Development, 61, 1697-1713.

Wickelgren, W. (1977). Speed-accuracy trade-off and information processing dynamics. Acta Psychologica, 41, 67-85.

Wilson, B. A., Evans, J. J., Alderman, N., Burgess, P. W. & Emslie, H. (1997). Behavioural assessment of the dysexecutive syndrome. In P. Rabbitt (Ed.), Methodology of Frontal and Executive Function. Hove: Psychology Press.